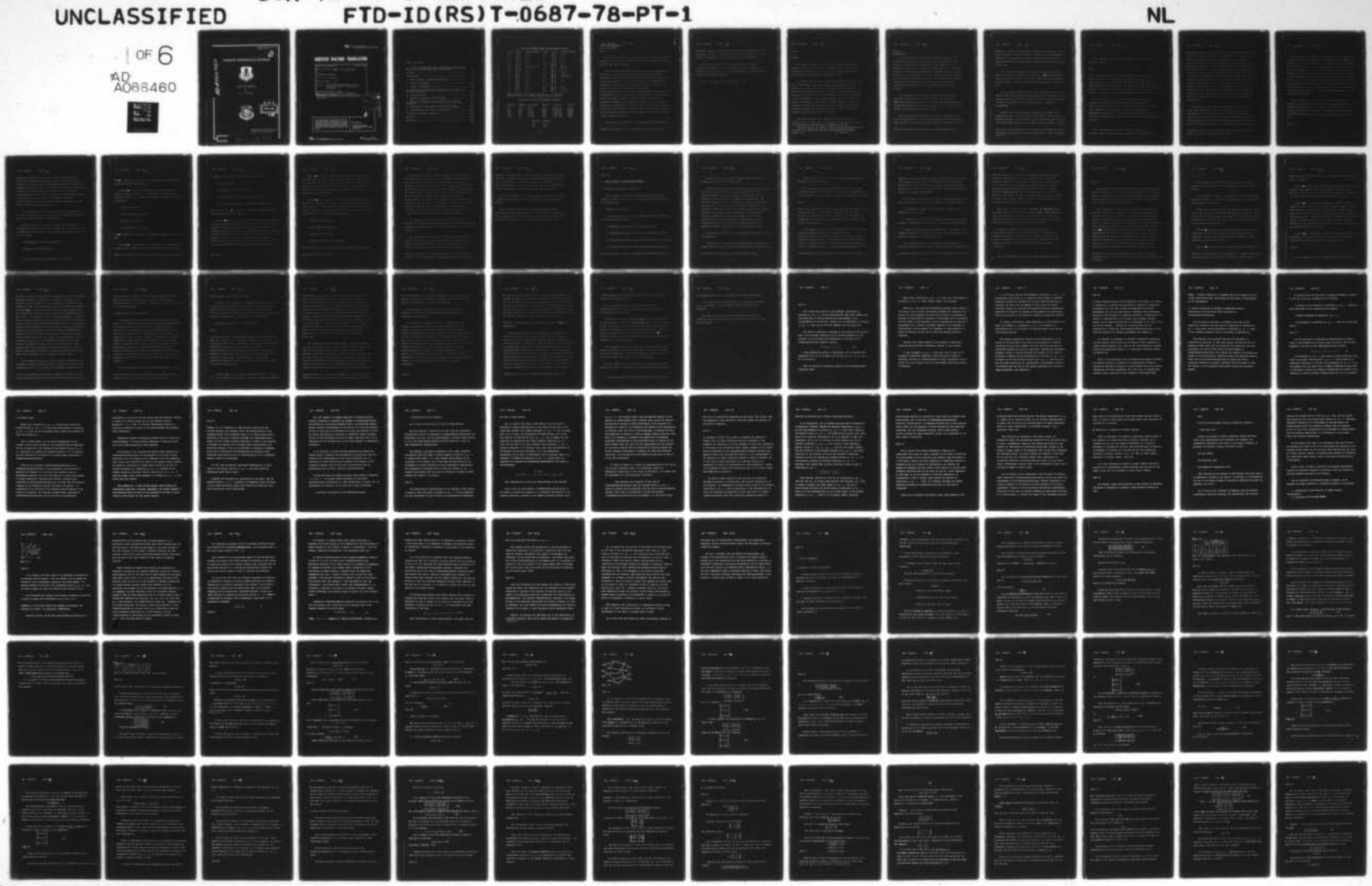
AD-A066 460 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/2
OPERATIONS RESEARCH. PART I,(U)
JUN 78 Y S VENTTSEL'
UNCLASSIFIED FTD-ID(RS)T-0687-78-PT-1 NL

1 OF 6
AD
A066460



FTD-ID(RS)T-0687-78
PART 1 of 3

AD-A066460

FOREIGN TECHNOLOGY DIVISION

AIR FORCE SYSTEMS COMMAND

OPERATIONS RESEARCH

By

Ye. S. Venttsel'

FOREIGN TECHNOLOGY DIVISION

Approved for public release;
distribution unlimited.

FTD- ID(RS)T-0687-78

UNEDITED MACHINE TRANSLATION

FTD-ID(RS)T-0687-78 8 June 1978

MICROFICHE NR: FTD-78-C-000783

OPERATIONS RESEARCH

By: Ye. S. Venttsel'

English pages: 1235

Source: Issledovaniye Operatsiy, Izd-vo "Sovetskoye Radio," Moscow, 1972, 546 pages.

Country of origin: USSR
This document is a machine translation.
Requester: FTD/TQFO
Approved for public release; distribution unlimited.

ACCESSION for	
NTIS	White Section ✓
DDC	Buff Section ☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

PREPARED BY:

TRANSLATION DIVISION
FOREIGN TECHNOLOGY DIVISION
WP-AFB, OHIO.

Table of Contents

U. S. BOARD ON GEOGRAPHIC NAMES TRANSLITERATION SYSTEM

Block	Italic	Transliteration	Block	Italic	Transliteration
А а	*А а*	A, a	Р р	*Р р*	R, r
Б б	*Б б*	B, b	С с	*С с*	S, s
В в	*В в*	V, v	Т т	*Т т*	T, t
Г г	*Г г*	G, g	У у	*У у*	U, u
Д д	*Д д*	D, d	Ф ф	*Ф ф*	F, f
Е е	*Е е*	Ye, ye; E, e*	Х х	*Х х*	Kh, kh
Ж ж	*Ж ж*	Zh, zh	Ц ц	*Ц ц*	Ts, ts
З з	*З з*	Z, z	Ч ч	*Ч ч*	Ch, ch
И и	*И и*	I, i	Ш ш	*Ш ш*	Sh, sh
Й й	*Й й*	Y, y	Щ щ	*Щ щ*	Shch, shch
К к	*К к*	K, k	Ъ ъ	*Ъ ъ*	"
Л л	*Л л*	L, l	Ы ы	*Ы ы*	Y, y
М м	*М м*	M, m	Ь ь	*Ь ь*	'
Н н	*Н н*	N, n	Э э	*Э э*	E, e
О о	*О о*	O, o	Ю ю	*Ю ю*	Yu, yu
П п	*П п*	P, p	Я я	*Я я*	Ya, ya

*ye initially, after vowels, and after ъ, ь; e elsewhere.
When written as ё in Russian, transliterate as yё or ё.

RUSSIAN AND ENGLISH TRIGONOMETRIC FUNCTIONS

Russian	English	Russian	English	Russian	English
sin	sin	sh	sinh	arc sh	$\sinh^{-1}$
cos	cos	ch	cosh	arc ch	$\cosh^{-1}$
tg	tan	th	tanh	arc th	$\tanh^{-1}$
ctg	cot	cth	coth	arc cth	$\coth^{-1}$
sec	sec	sch	sech	arc sch	sech^{-1}
cosec	csc	csch	csch	arc csch	csch^{-1}

Russian	English
rot	curl
lg	log

OPERATIONS RESEARCH.

Ye. S. Venttsel'.

Page 2.

Are stated the bases of operations research - science, which is occupied by the quantitative proof of solutions in all fields of the goal-directed human activity.

In the book are examined basic concepts and the methodological principles of operations research, the mathematical methods of optimization (linear, dynamic programming, the theory of games and statistical solutions), and also the methods of the mathematical simulation of operations. Considerable attention is given to the applied theory of the Markovian processes (with application/appendices in the range of the queueing theory, theory of reliability) and to the mathematical description of the processes, which take place in complex, multiunit systems (method of the dynamics of average). Are examined the methods of the statistical simulation of operations by ETsVM [digital computer] and the bases of the method of statistical simulation. The book contains the number of the new materials, worked out by the author in recent years and than nowhere earlier published.

Presentation is conducted at a comparatively elementary level, completely available to reader, familiar with the usual higher

educational course of mathematics and with the cell/elements of the probability theory. The set-forth methods are illustrated by a large quantity of examples from the different ranges of practice.

The book is intended to the wide circle of the readers - engineers, economists, of the scientific workers and economic leaders, who are interested in the application/use of mathematics to the proof of optimum solutions.

Figs. - 266, Tables - 119, References - 29.

Page 6.

PREFACE.

This book is written on the basis of lectures on operations research, read by the author over a number of years in the higher educational institutions, and also on the basis of an experiment in the scientific research works in different fields.

By author's task was give as much as possible idle time and the clear presentation of ideas and methods of operations research, without using bulky mathematical apparatus. In the relation of mathematical preparation of reader is required only acquaintance with the usual VTUZ course of higher mathematics, and also the possession of the cell/elements of the probability theory. For the purpose of clarity presentation is accompanied by many examples. The book is intended to the wide circle of the readers - mainly, engineers and scientific workers, who are interested in the tasks of the proof of solutions in different ranges of practice.

The author expresses deep gratitude to I. Ya. Diner, L. A. Ovcharov and A. D. Venttsel', collaboration with which aided it in the development of the materials, presented in the book.

The book contains the number of the new materials, worked out be the author in recent years and than nowhere earlier published.

Moscow, 1970 Ye. Venttsel'.

Page 7.

INTRODUCTION.

In recent years the science pays increasing attention to questions of organization and control; this is caused by a whole series of reasons. Rapid development and the complication of technology; an increase of scales and cost/values of the conducted measures; the widespread introduction of automation into the sphere of control - all this leads to the need for the scientific analysis of the complex goal-directed processes at the visual angle of their structure and organization. Of science they are required for recommendation regarding the best (optimum) control of such processes.

These necessities of practice caused to life the special scientific methods which it is accepted to join by the name "Operations research". Under this is implied the application/use of mathematical, quantitative methods for the proof of solutions in all fields of the goal-directed human activity.

The need for making decisions so is old as humanity herself. From time immemorial to century people, beginning the realization of their measures, considered above their possible consequences and made

decisions, choosing in some way or other the depending on them

parameters - methods of organizing the measures. But until a certain time of solution we could be accepted without special mathematical analysis, it is simple on the basis of experiment and the common sense. This method of making decisions did not lose its value, also, in our time.

Let us take the example: man left by the morning to house in order to go to work. On course of events for it [redacted] is necessary to take a whole series of the solutions: to take with itself umbrella? In which place to pass street? Which form of transport to use? and so on.

It goes without saying that all these decisions of man are made without special calculations, simply relying on the available he has experiment and on the common sense. For the proof of such solutions, no science is necessary, yes scarcely it will be required subsequently.

However, let us take another example. Let us assume that is organized the work of urban transport. Available is some quantity of conveying devices. It is necessary to take a series of the solutions, for example: which quantity and of which conveying devices to direct along one or the other route? As to change the repetition frequency of machines depending on the time of days? Where to place cessations? and so on.

These solutions are much more critical, than the solution of the previous example.

Page 8.

Because of the complexity of the phenomenon of the consequence of each of them are not so/such clear; in order to visualize these consequences, it is necessary to lead calculations. But main, on these solutions much more depends. In the first example the incorrect selection of solution will affect the interests of one person; in the second - it can be reflected in the business life of whole city.

It is certain, and in the second example when selecting of solution it is possible to function intuitively, relying on experiment and the common sense. But solutions will render/show much more reasonable, if they will be supported by quantitative, mathematical calculations. These preliminary calculations will aid to avoid the prolonged and expensive search of the correct solution "hit or miss".

Most complicatedly is matter with the acceptance of the solutions when speech occurs about the measures, experiment in

conducting of which still does not exist and, to consequently common sense not on that to be based on, but intuition can deceive. Let, for example, be comprised the long-range plan of the development of weapon system on several years forward. The specimen/samples of the armaments about which can go the speech, still do not exist, there is no experiment in their combat employment. During gliding/planning it is necessary to be based on a large quantity of data, that relate not so much to past experiment, as to the foreseen future. The selected solution possibility to must as far as guarantee us from the errors, connected with inaccurate forecasting, and to be is sufficient efficient for the wide circle of conditions. For the proof of this solution, is given into action the complex system of mathematical calculations, yes otherwise and to be not must: indeed incorrect solution, if it will be accepted that it can lead to the heaviest consequences.

Generally, the more complexly organized measure, the more is packed into it supplies, the wider the spectrum of its possible consequences, the less permissible the so-called "volitional" solutions, which are not based on scientific calculation, and the larger value obtains the set of the scientific methods, which make it possible to previously consider the consequences of each solution, to previously reject/throw the inadmissible versions and to recommend those, that are represented most successful.

By such mathematical by the calculations, which facilitate for people making correct decisions, and is occupied science "Operations research". This - comparatively young science. Its emergence usually is related to the years of the Second World War when in armed forces of USA and England were formed special scientific groups for the preparation of solutions by the methods of organization and provision for combat operations [1]. Validity requires to note that by similar research (true, not under this name) they were occupied to the war, in particular, in our country where were widely developed the mathematical methods of the evaluation of the efficiency of shooting, representing by themselves, in contemporary understanding, the part of operations research.

After conceiving in the range of predominantly military problems, operations research in the course of time left this narrow sphere. At present operations research - one of the rapidly developing sciences themselves, the conquering ever faster fields of application industry, agriculture, trade, transport, public health, etc.

Page 9.

The tasks of operations research, to whatever range they were related, have common/general/total features, and during their solution are applied similar methodological methods. For example, the methodology of quantitative study, manufactured for the analysis of the processes of queueing in the systems of mass maintenance (barbershop, repair shops and so forth), can almost without changes, to be postponed by some tasks of electronics computational engineering, and also task, connected with the organization of air defense system (PVO [Air Defense]).

In order nearer to be introduced to the specific character of the tasks of operations research and their characteristic features, let us give several examples of such tasks.

Example 1. Plant is issued the specific type of article. For providing the high quality of these articles, is organized the system of sampling. It is required by rational form to organize this check, i.e., to select:

- size/dimension of control party/batch;

- sequence of control operations;

- rule of the rejection of spoilage of the articles

and [redacted] so forth so as to ensure the assigned level of quality with minimum expenditure/consumptions.

Example [redacted] 2. For the realization of the specific mass of seasonal goods, is created the grid/network of time/temporary commercial points. It is required to select the parameters of this grid/network:

- number of points;

- their arrangement/position;

- quantity of personnel;

- sellings-price of the goods

and [redacted] so forth so as to ensure the maximum economic efficiency of sale.

Example [redacted] 3. Is organized the air raid of the group of bomber aircraft in the industrial region of enemy. Available - specific

quantity of aircraft with known tactical flight data and armament. It

is required to select the parameters of the coating:

- flight altitude;

- separation of aircraft in system;

- aiming point of separate aircraft and groups;

- method of the execution of bombing (by volley, series)

and so forth so as a result of coating to maximally reduce the industrial potential of region.

Example 4. Is organized the supply with the raw material of the group of industrial enterprises. The possible suppliers of raw material are placed in different geographical point/items of the country and are connected with the group of the enterprises variously of report/communication (with different tariffs). It is required by rational form to place orders on raw material, so that the necessities of the group of enterprises would be satisfied for the assigned periods and with minimum expenditures on transport.

Page 10.

Example 5. Complex technical equipment/device from time to time can reject (go out of order). For eliminating of emergency, it is necessary to localize malfunction (to discover its reason). It is required to work out the system of tests, which makes it possible with the specific, sufficiently large probability to localize malfunction for minimum time.

Example 6. Is organized the medical examination/inspection of the group of population for purpose of the development/detection of some diseases. To examination/inspection are isolated the specific supplies, equipment and medical personnel. It is required to work out the plan/layout of the examination/inspection:

- quantity of point/items;

- their arrangement/position;

- sequence of inspections;

- form and a quantity of the analyses

and so forth in order to the assigned period to detect the maximum percentage of those sickened.

The number of examples possible would be easy to multiply, but also these are sufficient in order to comprise the representation of the distinctive special feature/peculiarities of the tasks of operations research. Despite the fact that examples are related to the different ranges of practice, in them easily are examine/scanned the similarities. In each of them occurs speech about some measure (or the system of measures), which pursues the specific target/purpose. Are assigned some conditions, which characterize the situation of the measures to change which we not right (for example, the tempered means). Within the framework of this system of conditions, is required to take some solution so that the measure in a sense would be most advantageous (or least unprofitable).

In accordance with these common/general/total features are developed the common/general/total methods of the solution of similar problems, in set which compose the methodological basis of operations research.

For the solution of practical problems, operations research disposes of the whole arsenal of mathematical means. To it are related: the probability theory with its newest sections (theory of random processes, information theory, queueing theory); the

mathematical methods of optimization, beginning from the easiest methods of the determination of extrema (maximums and the minimums),

familiar tc each engineer, and ending with the contemporary methods, with such, as linear programming, dynamic programming, the principle of L. S. Pontriagin's maximum and many cthers. From them in this bock, addressed to the wide circle of the readers, are illuminated by no means everything, but cnly simplest and most widely used.

For the understanding cf text, the reader must manage only bases cf mathematical analysis and the cell/elements of the probability theory.

In the book are contaired many numerical examples, which illustrate the set-forth methods. When selecting of the conditions of these examples, the author proceeded frcm systematic considerations, so that these materials in nc case cannot be used as reference.

Page 11.

1. BASIC CONCEPTS OF OPERATIONS RESEARCH.

1. Operation. Efficiency of operation.

Under operation we will understand any measure (or action system), united by single project and directed toward achieving of the specific goal.

Examples of operations.

1. System of measures, directed toward increase of reliability of technical equipment/device.

2. Reflection of air raid by air defense weapons.

3. Arrangement/position of orders to production of equipment.

4. Reconnaissance search of groups of aircraft in rear of enemy.

5. Starting/launching of group of artificial Earth satellites for establishment of system of television communication/connection.

6. System of transport, which ensures supply of series of point/items of specific form with goods.

Operation is always the controlled measure, i.e., on us depends to select in this or some other way some parameters of its characteristic method of organization. "Organization" here is understood in the broad sense, including the selection of the technical equipment, used in operation. For example, organizing the reflection of the air raid by the air defense weapons, we can, depending on situation, choose type and the properties of technical equipment used (rockets, installations) or, with the assigned technical equipment, to solve only problem of the rational organization of the very procedure of reflection it is put on (distribution of the target/purposes between installations, a quantity of rockets, directed to each target/purpose, etc.).

Any specific selection on us of the parameters depending we will call solution.

Solutions can be successful and unsuccessful, reasonable and unreasonable. Optimum are called the solutions which according to one or the other considerations, are more preferable others.

Basic task of operations research - preliminary quantitative foundation of optimum solutions.

Let us note that very making of decision exceeds the scope operations research and is related to the scope of responsible person (or the group of persons), which it is given rightly final selection.

Page 12.

During this selection the critical for it persons can consider that together with the recommendations, which escape/ensue from mathematical calculation, another a series of the considerations (quantitative and qualitative character) which were not taken into account by calculation.

Thus, operations research is not placed to itself with task the full/total/complete automation of making decisions, full/total/complete exception/elimination from this process of the reflecting, evaluating human consciousness. As the final result, solution is always received by man (or the group of persons); the task of operations research - to prepare quantitative data and the recommendations, which facilitate for man acceptance of solution [1].

FOOTNOTE [1]. Even when making decision, it would seem, fully automated (for example, in the process of automatic control enterprise or spacecraft), the role of man is not removed, since, in the final analysis, on it depends the selection of the algorithm, on which is realized the control. ENDFOOTNOTE.

Together with basic task - the proof of optimum solutions - the area of exploration of operations includes other problems, such as

- a comparative evaluation of the diverse variants of the organization of operation;

- evaluation of effect on the result of operation of different parameters (cell/elements of solution and the assigned conditions);

- study of the so-called "bottlenecks", i. e., elements of the controlled system the disruption of work of which especially strongly affects the success of operation and, etc.

These "auxiliary" tasks of operations research acquire the special importance when we examine this operation not isolatedly, but as component element of the whole system of operations. The so-called

"systems" approach to the tasks of operations research requires the account to interdependency and the conditicnality of the whole complex cf measures. It goes without saying that in principle always it is possible to join the system of operations into one complex operation more "high order", but in practice this not is always convenient (and it is not always desirable), and in a series of the cases it is expedient tc select as "operations" the separate elements cf system, but the final decision to make taking into account role and place of this operaticn in system.

Thus, let us consider separate operation O. Reflecting above the organization of operaticn, we attempt to make it most efficient. Under the efficiency of operaticn, is understood the degree of its adaptability to the accomplishment of the confronting it objective. The better is organized operation, the more efficient.

In order to judge the efficiency cf operation and to equate/compare between themselves with respect to efficiency the differently crganized operations, it is necessary to have certain numerical evaluaticn criteria or index cf the efficiency (in some management/manuals the index of efficiency calls "objective function").

Let us subsequently designate the index of efficiency by letter W.

Page 13.

The concrete/specific/actual form of the index of efficiency W, which one should use during the numerical evaluation of efficiency, depends on the specific character of the operation of its purposeful directionality in question, and also on the task of research which can be placed in one or the other form.

Many operations are made under conditions, which contain the element of chance (for example, the operations, connected with the fluctuations of supply and demand, with the motion of population, morbidity, mortality, and also all military operations). In these cases the issue of operation, even organized by the strictly defined form, it cannot be accurately predicted, remains random. If this then, [redacted] as the index of efficiency W is chosen not simply the characteristic of the issue of operation, but its average value (mathematical expectation). For example, if the task of operation - obtaining maximum gain, then as the index of efficiency is taken average gain. In other cases when the task of operation is the realization of the completely specific event, as the index of

efficiency, is taken the probability of this event (for example, probability that as a result of the air raid this target complex will be struck).

Correct selection of the index of efficiency - necessary condition of the usefulness of the research, used for the proof of solution.

Let us consider a series of the examples, in each of which the index of efficiency W is selected in accordance with the purposeful directionality of operation.

Example 1. Is examined the work of industrial enterprise at the visual angle of its profitableness, moreover is conducted the number of measures for target/purpose for purpose of the increase of this profitableness. Index of efficiency - gain (or average gain), yielded by enterprise for fiscal year.

Example [redacted] 2. The group of destroyers heaves into air for the interception of the single aircraft of enemy. Target/purpose of operation - to bring down aircraft. Index of efficiency - kill probability (killing) of aircraft.

Example [redacted] 3. Repair shop is occupied by the maintenance of machines; its profitableness is determined by a quantity of machines, serviced during day. The index of efficiency - average number of

machines, serviced for the day (" it is average" because the actual number is random).

Example 4. The group of radar stations in the specific region observes after airspace. The task of group - to discover any aircraft, if it appears in region. Index of efficiency - probability of the detection of any aircraft, which appeared in region.

Example 5. Is launched the number of measures for the increase of the reliability of electronic digital computer (ETsVM). Target/purpose of operation - to decrease the frequency of the appearance of malfunctions ("short duration failures") of ETsVM, or, that equivalently, to increase average time interval between short duration failures (the "mean time between failures"). Index of efficiency - mean time of the failure-free operation of ETsVM (or the mean relative time of exact work).

Example 6. Is conducted fight for the economy of means in the production of the specific form of goods. Index of efficiency - quantity (or an average quantity) of economized means.

Page 14.

In all examined examples the index of efficiency, whatever it

- rule of the rejection of spoilage of the articles

was, it was required to convert into maximum ("the more the better"). Generally, this not is compulsory: in operations research, they frequently use the indices which it is required to convert not into maximum, but into the minimum ("the fewer, the better"). For example, in example 4 it would be possible as the index of efficiency to take "probability that the appearing aircraft will not be discovered" - this index it is desirable to make as small as possible. In example 5 for the index of efficiency it would be possible to take the "average number of short duration failures for days", which it is desirable to minimize. If is considered some system, which ensures the guidance of projectile to target/purpose, then as the index of efficiency it is possible to select the average value of the "error" of projectile (distance from trajectory to the center of target/purpose), which it is desirable to do as less as possible. The detail of the means, isolated on carrying out any task, it is also desirable to make minimum, just as the cost/value of the launched system of measures. Thus, in many tasks of operations research reasonable solution must ensure not the maximum, but the minimum of certain index.

It is obvious that the case, when the index of efficiency W must be converted into the minimum, easily is reduced to the task of the maximization (for this it suffices, for example, to change the sign of value W). Therefore subsequently, examining in general form the task of operations research, we will for simplicity speak only about

the case when W is required to convert into maximum. As concerns practical specific problems, then we will use as indices of efficiencies which it is required to maximize, so also those which it is required to minimize.

2. Mathematical model of operation.

For applying the quantitative methods of study in any field, always it is required to construct one or another mathematical model of phenomenon. Operations research is no exception. During the construction of mathematical model the phenomenon (in our case - operation) by some form it is simplified, it is schematized; from numerous set of factors, which affect the phenomenon, is selected a comparatively small quantity of the important, and the obtained pattern is described with the help of one or the other mathematical apparatus. As a result are establish/installed quantitative communication/connections between the conditions of operation, the parameters of solution and the issue of operation - an index of efficiency (or by indices, if there are several of them in this problem).

The more successful is selected mathematical model, the better

it reflects the characteristic features of phenomenon, the more successful will be research and it is more useful - the

escape/ensuing from it recommendations.

Methods of the construction of mathematical models do not exist. In each specific case model is constructed, on the basis of the purposeful directionality of operation and task of scientific research, taking into account the required accuracy of solution, also accuracy, from which there can be known the initial data.

Requirements for model are contradictory.

Page 15.

On one hand, it must be sufficiently complete, i.e., in it must be taken into account all the important factors on which depends substantially the issue of operation. On the other hand, model must be sufficient simple so that it would be possible to establish/install the foreseeable (are desirable - analytical) dependences between the entering it parameters. Model must not be "clogged" by many small, secondary factors - their account complicates mathematical analysis and makes the results of research by difficultly foreseeable.

In a Word, the art to comprise mathematical models is precisely art, and experiment in this matter is acquired gradually. Two dangers

are always on the watch the compiler of the model: the first - to drown in the details ("can't see the forest for the trees"); the second - to too desensitize the phenomenon ("to throw the baby out with the bath water"). In the complex cases when the construction of model causes great doubt, useful proves to be the peculiar "dispute of models", when one and the same phenomenon is traced on several models. If scientific conclusion/derivations and recommendations from one model to the next vary barely, this - serious argument in favor of the objectivity of research. Characteristic for the complex problems of operations research is also repeated handling to the model: after the first cycle of research it is carried out, are returned again to model and introduce into it the necessary corrections.

The construction of mathematical model - the most important and critical part of the research, which requires intimate knowledge not only and not so much in mathematics, as in the essence of the phenomena being simulated. However, once the created successful model can find use and far beyond the limits of that circle of phenomena, for which it initially was created. Thus, for instance, mathematical models of mass maintenance had extensive application in a whole series of maintenance (reliability of technical equipment/devices,

the organization of the automated production, task PVO, etc.). The mathematical models, initially intended for describing the dynamics

of the development of biological populations, find wide application during the description of combat operations and vice versa - combat models successfully they are applied in biology.

The mathematical models, used at present in the tasks of operations research, can be roughly subdivided into two class: analytical and statistical.

For the first is characteristic the establishment of the formula, analytical dependences between the parameters of the problem, registered in any form: algebraic equations, ordinary differential equations, partial differential equations and so forth. So that this analytical description of operation would be possible, as a rule, it is necessary to take these or other assumptions or simplifications. With the help of analytical models with satisfactory accuracy it is possible to describe only comparatively simple operations where the number of interacting cell/elements not is too great. In the operations of large scale, are complex, in which is interwoven the action of an enormous quantity of factors, including random, to the foreground leaves the method of statistical simulation.

Page 16.

It lies in the fact that the process of operations development as "is copied" in computer, with its all accompanying chances. Every time that in the course of operation is add/interfered any random factor, its effect is considered by the means of the "drawing", which recalls the throwing of toss. As a result of the multiple repetition of this procedure succeeds in obtaining these interest us the characteristics of the issue of operation with any degree of accuracy [1].

FOOTNOTE [1]. About statistical simulation see in detail Chapter 8. ENDFOOTNOTE.

Statistical models have before analytical the advantage that they make it possible to take into account the larger number of factors and do not require rough simplifications and assumptions. Then the results of statistical simulation with more difficulty yield to analysis and comprehension. Rougher analytical models describe phenomenon only approximately, then results are more demonstrative and more distinct reflect those be inherent in phenomenon basic laws. Best results are obtained during the combined application/use of the analytical and statistical models: simple analytical model makes it possible to be dismantle/selected in the rough at the basic laws governing phenomenon, to plan its main outlines, and any further

refinement can it can be obtained by statistical simulation.

3. Common/general/total formulation of the problem of operations research the determined case.

Let us consider the task of operations research in common/general/total setting, regardless of the form and the target/purposes of operation.

Let there be certain operation **O, i.e.,** the controlled measure to issue of which we can to a certain degree affect, choosing in this or some other way on us the parameters depending. The efficiency of operation is characterized by some numerical criterion or index W, which is required to convert into maximum (case when it is required to convert into the minimum, it is reduced to previous and separately it is not examined).

Let us assume that in one or the other manner the mathematical model of operation is constructed; it makes it possible to compute the index of efficiency W during any solution accepted, for any set of the conditions, under which is made the operation.

Let us consider the first simplest case: all factors on which depends the success of operation, they are divided into two groups:

- assigned, previously known factors (conditions of conducting

the operation) a_1, a_2, ..., over which we can have no influence;

- on us the factors (cell/elements of the solutions) depending x_1, x_2, ..., which we, within known limits, can choose at our discretion.

This case, in which the factors, influencing the issue of operation, either are previously known or they depend on us, we will call determined.

Page 17.

Let us note that hearth by the "assigned conditions" of operation $\alpha_1, \alpha_2, \ldots,$ can be understood not only usual numbers, but also functions, in detail-limitations, superimposed to the cell/elements of the solution. Equally, the cell/elements of solution $x_1, x_2, \ldots,$ also can be not only numbers, but also functions.

The index of efficiency W depends on both groups of the factors: both on the assigned conditions and on the cell/elements of the solution. Let us register this dependence in the form of the common/general/total symbolic formula:

$$W = W(\alpha_1, \alpha_2, \ldots; x_1, x_2, \ldots). \qquad (3.1)$$

Since mathematical model is constructed, let us consider that dependence (3.1) to us is known, and for any $\alpha_1, \alpha_2, \ldots; x_1, x_2, \ldots$ we we can find W.

Then the problem of operations research can be mathematically formulated thus:

Under given conditions α_1, α_2, ... to find such cell/elements of solutions x_1, x_2, ..., which convert index W into maximum.

Before us - the typically mathematical problem, which relates to the class of the so-called variational problems. The methods of the solution of such problems are worked out in detail in mathematics. Protozoa of these methods ("problem to the maximum and the minimum") are well known to each engineer. For the determination of maximum or the minimum (it is shorter, extremum) function it is necessary to differentiate it the argument (or arguments, if them several), to equate derivatives to zero and to solve the obtained system of equations.

However, this simple method in the problems of operations research being of limited usefulness. Reasons to this several.

1. When arguments x_1, x_2, ... much (but this is typical for problems of operations research), joint solution of system of equations, obtained by differentiation of basic dependence, often it proves to be not simpler, but it is more complex than direct search of extremum.

2. In the case when to cell/elements of solution x_1, x_2, ... are superimposed limitations (i.e., region of their change is limited), frequently extremum is observed not at point where derivatives are converted into zero, but on boundary of the region of possible solutions. Appears the specific for operations research mathematical problem of the "search of extremum in the presence of limitations", which is not placed in the diagram of classical variational methods.

3. Finally, derivatives, under discussion, it can completely not exist, for example, if arguments x_1, x_2, ... are changed not continuously, but it is discrete, or function itself W has special feature/peculiarities.

The general mathematical methods of the determination of the extrema of the functions of any form in the presence of arbitrary limitations do not exist. However, for the cases when function and limitations possess the specific properties, contemporary mathematics proposes a series of special methods. For example, if the index of efficiency W depends on the cell/elements of solution x_1, x_2, ... it is linear and the restrictions placed on x_1, x_2, ..., also take the form of linear equalities (or inequalities), the maximum of function W is located with the help of the special apparatus, the so-called linear programming (see Chapter 2).

Page 18.

If these functions possess other properties (for example, are convex or quadratic), is applied the apparatus of "convex" or "quadratic" programming [2], more more complex in comparison with linear programming, but all the same making it possible within acceptable periods to find solution. If operation naturally is dismembered to a series of "step/pitches" or of "stages" (for example, economic years), and an index of efficiency W is expressed in the form of the sum of the indices w_i reached for various stages, for the determination of solution, which ensures maximum efficiency, can be used the method of the dynamic programming (see Chapter 3).

If operation is described by ordinary differential equations, and the control, which varies in the course of time, represents by itself certain function x(t), then for the determination of optimum control can render/show useful L. S. Pontryatin's specially worked out method [3].

Thus, in the determined case in question the problem of finding the optimum solution is reduced to the mathematical problem of finding the extremum of function W; this problem can be very complex (especially with many arguments), but, after all, is computational problem, which, especially in the presence of high speed ETsVM

[ЭЦВМ - digital computer], it succeeds one way or another to solve to end. The difficulties, which appear in this case, are calculated, but not fundamental.

4. General formulation of problem of operations research. Optimization of the solution under conditions of indeterminancy/uncertainty.

In the previous paragraph we considered idle time itself, completely determined the case when all conditions of operation α_1, α_2, ... were known, and any selection of solution x_1, x_2, ... leads to the completely specific value of the index of efficiency W.

Unfortunately, this simplest case not too frequently is encountered in practice. Is much more typical the case when not all conditions under which will be carried out the operation, are known previously, but some of them contain the cell/element of indeterminancy/uncertainty. For example, the success of operation can depend on the meteorological conditions which are previously unknown, either from the oscillation/vibrations of supply and demand, previously difficultly foreseen, connected with whims Maud, or from the behavior of the reasonable enemy whose actions are previously unknown.

In similar cases the efficiency of operation depends no longer on two, but from three categories of the factors:

- condition of the execution of operation α_1, α_2, ... which are known previously and are changed be they cannot;

- unknown conditions or factors Y_1, Y_2, ...;

- cell/elements of solutions x_1, x_2, ..., which for us one must select.

Page 19.

Let the efficiency of operation be characterized by certain index W, which depend on all three groups of factors. This we will register in the form of the common/general/total formula:

$$W = W(\alpha_1, \alpha_2, \ldots; Y_1, Y_2, \ldots; x_1, x_2, \ldots). \qquad (4.1)$$

If conditions Y_1, Y_2, ... were known, we could previously count index W and select such solution x_1, x_2, ..., during which it is maximized. Misfortune in the fact that parameters Y_1, Y_2, ... to us are unknown, and that means that is unknown depending on them index of efficiency W during any solution. Nevertheless the problem of the selection of solution stands as before before us. It it is possible

to formulate thus:

Under given conditions $\alpha_1, \alpha_2, \ldots$, taking into account the unknown factors $Y_1, Y_2, \ldots$ to find such cell/elements of solutions $x_1, x_2, \ldots$, which as far as possible would convert into maximum the index of efficiency W.

This - already other, not the purely mathematical of the problems (is not without reason in its formulation done stipulation "as far as possible"). The presence of the unknown factors $Y_1, Y_2, \ldots$ translates our problem into another category it it is converted into the problem of the selection of solution under conditions of indeterminancy/uncertainty.

Give let us be honest: indeterminancy/uncertainty is an indeterminancy/uncertainty. If the conditions of the execution of operation are unknown, we do not have the capability then to successfully organize it, as we this would do, if they would dispose of larger information. Therefore any solution, accepted under conditions of indeterminancy/uncertainty, is worse than the solution, accepted in the completely specific situation. Our matter - to communicate to its solution in the greatest possible measure of the feature of soundness. The solution, accepted under conditions of indeterminancy/uncertainty, but on the basis of mathematical

calculations, will be all the same better than the solution, selected at random. Not without reason one of the prominent foreign specialists - T. L. Saati in the book "mathematical methods of operations research" [4] gives to its object/subject the following ironic definition:

"Operations research represents by itself the art to give poor answer/responses to those practical questions, to which are given even worse answer/responses by other methods".

The problems of the selection of solution under conditions of indeterminancy/uncertainty are encountered to us in life at each step/pitch. Let, for example, we be collected to go into tempering, after taking with ourselves the trunk of the limited volume, moreover the weight of trunk must not exceed that by which we can bear it without extraneous aid (condition a_1, a_2, ...). Weather in the journey regions is previously unknown (condition Y_1, Y_2, ...). It does ask itself, which object/subjects of clothing (x_1, x_2, ...) one should take with itself?

This problem we, it goes without saying, solve without any mathematical apparatus, although, apparently, not without support to some numerical data (at least on the probability of frost or rainy weather in)the journey in this season regions.

Page 20.

However, if it is necessary to make the more serious and more critical decision (for example, about the characteristics of the design/projected dam in possible region it is flood, either about a selection of the type of landing equipment for landing/fitting on planet with the unknown surface properties, or about of the sampling of armament for dealing with the enemy whose characteristics are previously unknown), then to the selection of solution in necessary order must be presupposed the mathematical calculations, which facilitate this selection and the communicating to it, in available measure, features of soundness.

In this case the methods used depend substantially on which nature of the unknown factors Y_1, Y_2, ... and which tentative information about them we avail.

Simplest and favorable for calculations is the case, when the unknown factors Y_1, Y_2, ... represent by themselves random variables (or the random functions), about which there are statistical data, which characterize their distribution.

Let, for example, we examine operation of railroad sorting station, attempting to optimize the process of maintain/serviceing the arriving to this station freight trains. Are previously unknown neither precise torque/moments of the arrival of trains nor quantity of cars in each train nor addresses with which are directed the cars. All these characteristics represent by themselves random variables, the law of the distribution of each of which (and their set) can be determined by the available data by the usual methods of mathematical statistics.

It is analogous, in each military operation are present the random factors, connected with scattering of projectiles, with the chance of the torque/moments of the target detection, etc. In principle all these factors can be studied by the methods of the probability theory, and for them can be obtained the laws of distribution (or, at least, numerical characteristics).

In the case when the unknown factors, which figure in operation - Y_1, Y_2, ... - are usual random variables (or the random functions) whose distribution, at least tentatively, is known, for the optimization of solution, can be used one of the two methods:

- artificial information to the determined diagram;

- "optimization on the average".

Let us pause in more detail at each of these methods.

The first method is reduced to the fact that the indefinite, probabilistic picture of phenomenon is approximately substituted that determined. For this, all the participating in problem random factors Y_1, Y_2,: are approximately substituted not random (as a rule, by their mathematical expectations).

This method is applied to advantage in the rough, tentative calculations when the range of random changes in values Y_1, Y_2, ... is comparatively small, i.e., they without large tension can be considered as not random. Let us note that the same method of the replacement of random variables by their mathematical expectations can successfully be applied also in cases when values Y_1, Y_2, ... possess large spread, but the index of efficiency W depends on them linearly (or almost it is linear).

Page 21.

The second method ("optimization on the average"), more complex, is applied, when the chance of values Y_1, Y_2, ... is very essential and the replacement of each of them by its mathematical expectation

can lead to large errors.

Let us consider this case in more detail. Let the index of efficiency W depend substantially on the random factors (let us for simplicity consider them random variables) Y_1, Y_2, ...; let us assume that to us is known the distribution of these factors, let us say, that density of distribution $f(y_1, y_2, \ldots)$. Let us assume that the operation is implemented many times, moreover conditions Y_1, Y_2, ... vary from one time to the next randomly. Which solution x_1, x_2, ..., should be selected? It is obvious, then, with which the operation on the average will be most efficient, i.e., the mathematical expectation of the index of efficiency W will be maximal. Thus, it is necessary to choose such solution x_1, x_2, ..., during which is converted into maximum the mathematical expectation of the index of the efficiency:

$$\overline{W} = M[W] =$$

$$= \iint \ldots \int W(\alpha_1, \alpha_2, \ldots; y_1, y_2, \ldots; x_2, x_2 \ldots) f(y_1, y_2, \ldots) dy_1 dy_2 \ldots \quad (4.2)$$

This optimization we will call "optimization on the average".

But as with the cell/element of indeterminancy/uncertainty? It is certain, in which that measure it is retained. The success of each separate operation, realized at the random, previously unknown values

Y_1, Y_2, ..., can strongly differ from the expected average as into greater, so, unfortunately and to smaller side. During the repeated realization of operation, these differences, on the average, are smoothed out; however, the frequently this method of the optimization of solution, after the lack of the better/best, is applied and then, when operation is realized a total of several times or even one time. Then it is necessary to consider the possibility of unpleasant unexpected contingencies in each individual case. As comfort us can serve thought about the fact that the "optimization on the average" all the same is better than the selection of solution without any proofs. Applying this method to numerous (at least and different) operations, all the same we on the average win more than if they in no way used calculation.

In order to comprise to itself the representation of that how we risk in each individual case, it is desirable, except the mathematical expectation of the index of efficiency, to consider also its dispersion (or root-mean-square deviation).

Most difficult for research is that case of indeterminancy/uncertainty when the unknown factors Y_1, Y_2, ... cannot be studied and are described with the help of the statistical methods: their laws of distribution or cannot be obtained (corresponding statistical data are absent), or, that still worse,

such laws of distribution completely do not exist. This occurs, when the phenomenon, under discussion, does not possess the property of statistical regularity.

Page 22.

For example, we know, that on Mars is possible the presence of organic life, and some scientists even are considered it very probable, but completely it is not possible to count this probability on the basis of any statistical data. Another example: let us assume that the efficiency of the design/projected armament strongly does depend on that, will the predicted enemy up to the torque/moment of the beginning of combat operations dispose of means of defense, and if yes, then by which precisely? It is obvious, there is no possibility to count the probabilities of these hypotheses - larger, it is possible to assign them arbitrarily which will strongly damage the objectivity of research.

In similar cases, instead of the arbitrary and subjective designation/purpose of probabilities with further "optimization on the average", is recommended to consider entire range of the possible conditions Y_1, Y_2, ... and the comprising of the representation of which the efficiency of operation in this range and as it affect unknown conditions. With this problem of operations research it

acquires new methodological special feature/peculiarities.

It is real/actual, let us consider the case when the efficiency of operation W depends, besides the assigned conditions α_1, α_2, ... and the cell/elements of solution x_1, x_2, ..., even from a series of the unknown factors Y_1, Y_1, ... the nonstatistical nature about which there are no specific information, and it is possible to make only assumption. Let us try all the same to solve problem. Let us fix mentally parameters Y_1, Y_2, ..., let us give to them completely specific values $Y_1 = y_1$, $Y_2 = y_2$, ..., and let us transfer thereby into the category of the assigned conditions α_1, α_2, ... For these conditions we in principle can solve the problem of operations research and find the appropriate optimum solution x_2, x_2, ... Its cell/elements, besides the assigned conditions α_1, α_2, ..., obviously, will depend even on which particular values we gave to conditions Y_1, Y_2, ...

$$x_1 = x_1(\alpha_1, \alpha_2, \ldots; y_1, y_2, \ldots);$$
$$x_2 = x_2(\alpha_1, \alpha_2, \ldots; y_1, y_2, \ldots).$$

This solution, optimum for this set of conditions y_1, y_2, ... (and only for it), is called local-optimum. This solution, as a rule, no longer is optimal for other values Y_1, Y_2, ... The set of local-optimum solutions for entire range of conditions Y_1, Y_2, ... gives to us the representation of how we must enter, if the unknown conditions Y_1, Y_2, ... were to us in accuracy known. Therefore

local-optimum solution on obtaining of which often are expended many effort/forces, has in the case of indeterminancy/uncertainty especially limited value. Is completely obvious that in this case one should prefer not the solution, strictly optimum for some specified conditions, but the compromise solution, which, being perhaps strictly optimum for which conditions, proves to be acceptable in the whole range of conditions.

Page 23.

The at present full-valued mathematical "theories of a compromise" still does not exist, although in the theory of solutions are some attempts in this direction (for example, see § 13 chapter of 9 this books). Usually the final selection of compromise solution is realized by man who, relying on calculations, can consider and compare the powerful and weak sides of each version of solution under different conditions also on the basis of this do a final selection. In this case, it is not necessarily (although sometimes it is interesting) to know a precise local optimum for each set of conditions y_1, y_2, ... Thus, the classical variation and newest optimization methods of mathematics step back in this case to background.

Lastly let us consider the peculiar case, which appears in the

so-called conflicting situations when the unknown parameters Y_1, Y_2, ... depend not on objective facts, but on actively counteracting to us enemy. Such situations are characteristic for combat operations, partly for sport competitions, in capitalist society - for a concurrent fight, etc.

When selecting of solutions in the similar cases, can render/show useful the mathematical apparatus of the so-called theory of games - mathematical theory of the conflicting situations (see Chapter 10). The models of conflicting situations, studied in the theory of games, based on the assumption that we deal with reasonable and farsighted enemy, who always chooses his behavior in the worst for us (and best for ourselves) manner. This idealization of conflicting situation in certain cases can prompt to us the least risky, "reinsurance" solution which to not necessarily accept, but in any case is useful to keep in mind.

Let us finally do one a generality. With the proof of solution under conditions of indeterminancy/uncertainty, that we not made, the cell/element of indeterminancy/uncertainty remains. Therefore it is unwise to present to the accuracy of such solutions of too high a requirement. Instead of after scrupulous calculations unambiguously indicating of only one, in accuracy optimum (in some sense) solution, it is always better to isolate the domain of the acceptable solutions

which prove to be unessentially worse than others, whatever point we used. In limits of this domain can produce their final selection the critical for it persons.

5. Estimation of operation by several indices.

Above we considered the problem of operations research where it was required so to select solution in order to maximize (or to minimize) the one and only index of efficiency W. In practice frequently is encountered the case, when the efficiency of operation is necessary to consider not according to one, but immediately according to several indices: $W_1, W_2, \ldots, W_k$; some of these indices desirable to do more, others - are less.

As a rule, efficiency of large in volume, complex operations cannot be in a comprehensive manner described with the help of one index; to aid for it is necessary to draw others, further.

Page 24.

For example, during the estimation of the activity of industrial enterprise is necessary to consider a whole series of indices as that:

-gain,

-the full/total/complete volume of production ("shaft"),

- prime cost, etc.

During the analysis of combat operation, besides the basic index, which characterizes its efficiency (for example, the mathematical expectation that caused to the enemy of damage), is necessary to consider a series of further as that:

-its own losses,

-the operation time,

-the ammunition consumption, etc.

This multitude of the indices of efficiencies from which some it is desirable to maximize, and others - to minimize, is characteristic for any of any complex problem of operations research. Does arise the question: how to be?

It is first of all necessary to emphasize that the advanced requirements, generally speaking, are incompatible. The solution,

turning into maximum some one index W_1, as a rule, converts either into maximum or into the minimum other indices W_2, W_3, ... Therefore widespread formulation the "achievement of maximum effect with minimum expenditures" for scientific research does not approach. Correct is any of the formulations the "achievement of maximum effect with the assigned expenditures" or "achievement of the assigned effect with minimum expenditures".

In the general case there is no solutions, which would convert into maximum one index W_1 and it is simultaneous into maximum (or the minimum) another index W_2; all the more, this solution does not exist for several indices. However, the quantitative analysis of efficiency can render/show highly useful, also, in the case of several indices of efficiency.

First of all, it makes it possible to previously reject/throw clearly the irrational versions of solutions, which are inferior to the best versions according to all indices.

Let us illustrate the aforesaid based on example. Let be analyzed the combat operation O, evaluated according to two indices:

W - probability of the execution of combat mission "efficiency");

S - cost/value of the spent means.

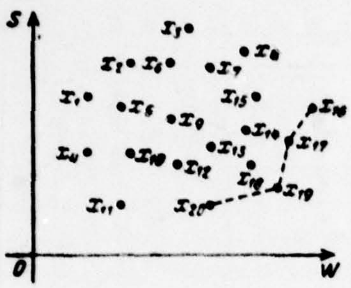


Fig. 1.1.

Page 25.

It is obvious, the first index it is desirable to convert into the maximum, and the second - into the minimum. Let us assume for simplicity that is proposed to selection the finite number - 20 diverse variants of solution; let us designate them X_1, X_2, ..., X_{20}. For each of them, are known the values of both indices W and S.

Let us describe for clarity each version of solution in the form of point on plane with coordinates W and S (Fig. 1.1) [1].

FOOTNOTE [1]. In the book figures are numbered on chapters, and formulas and tables - on paragraphs. ENDFOOTNOTE.

Examining figure, we see that some versions of solution "are

noncompetitive" and previously must be reject/thrown. It is real/actual, those versions which have above other versions with the same cost/value S advantage on efficiency W, they must lie/rest on the right boundary of the region of possible versions. The same versions, which with an equal efficiency possess smaller cost/value, must lie/rest on the lower boundary of the region of possible versions.

Which versions one should prefer during the evaluation of efficiency according to two indices? Obviously, those that lie/rest simultaneously both on the right and on lower boundary of the region (see dotted line in Fig. 1-1). It is real/actual, for each of the versions, which do not lie on this section of boundary, always will be located another version, which is not inferior to it on efficiency, but cheaper or, on the contrary, not being inferior to it on cheapness, but more efficient. Thus, of 20 advanced versions majority drops out from competition, and to us there remains only to analyze the remaining four versions: X_{16}, X_{17}, X_{19}, X_{20}. From them X_{16} - most efficient, but comparatively by road; X_{20} - cheapest, but not so/such efficient. The matter of that making decision - to be dismantle/selected at by which value we are concordant to pay the known increase of efficiency or, on the contrary, by which portion/fraction of efficiency we are concordant to endow in order not to carry too great material losses.

The analogous preliminary survey of versions (although without this demonstrative geometric interpretation) can be produced also in the case of many indices: $W_1, W_2, \ldots, W_k$.

This procedure of the preliminary reject of the noncompetitive versions of solution must always precede the solution of the problem of operations research with several indices. This, although does not remove/take the need for a compromise, substantially reduces the solution set within limits of which is realized the selection.

In view of the fact that the composite estimation of operation is immediately according to several indices difficult and requires speculations, in practice frequently they attempt to artificially join several indices into one generalized index (or criterion). Frequently as this generalized (compound/composite) criterion is taken fraction; in numerator place those indices $W_1, \ldots, W_m$, which it is desirable to increase, and in denominator, those which it is desirable to decrease:

$$U = \frac{W_1 \ldots W_m}{W_{m+1} \ldots W_k}. \qquad (5.1)$$

Page 26.

For example, if speech occurs about combat operation, in numerator place such values as "the probability of the fulfillment of combat mission" or the "losses of enemy"; in denominator - "his own losses", "ammunition consumption", the "operation time", etc.

An overall deficiency/lack in the "compound/composite criteria" of type (5.1) it is, the fact that a deficiency/lack in the efficiency according to one index always it is possible to compensate for because of another (for example, small probability of the fulfillment of combat mission - because of the low ammunition consumption and, etc.). Similar criteria call to mind into joke proposed to Leo Tolstoy "evaluation criteria of man" in the form of the fraction where the numerator - true advantages of man, and denominator - its opinion about itself. The groundlessness of this criterion is obvious: if we take it in earnest, then man, almost without advantages, but entirely without conceit, will have infinite value !

Frequently "compound/composite criteria" are proposed not in the form of fraction, but in the form of the "weighted sum" of the separate indices of the efficiency:

$$U = a_1 W_1 + a_2 W_2 + \ldots + a_k W_k, \qquad (5.2)$$

where $a_1, a_2, \ldots, a_k$ - positive or negative coefficients. Positive are

placed with those indices which it is desirable to maximize; negative - with those which it is desirable to minimize. The absolute values of coefficients ("weight") correspond to the degree of the importance of indices.

It is not difficult to ascertain that the compound/composite criterion of form (5.2) actually in no way differs from the criterion of form (5.1) and possesses the same deficiency/lacks (possibility of the mutual compensation heterogeneous indices). Therefore the noncritical use of any form of "compound/composite" criteria fraught with dangers can lead to incorrect recommendations. However, in certain cases when "weights" are not chosen arbitrarily, and they are selected so that the compound/composite criterion would implement in the best way its function, it is possible to obtain with its aid some results of the limited value.

In certain cases problem with several indices can be reduced to problem with some-only index, if we isclate only one (main) index of efficiency W_1 and to attempt it to convert into maximum, but to the remaining, auxiliary indices W_2, W_3, ... to superimpose only some limitations of the form:

$$W_2 \geqslant w_2; \ldots; \quad W_m \geqslant w_m; \quad W_{m+1} \leqslant w_{m+1}; \ldots; \quad W_k \leqslant w_k.$$

These limitations, it goes without saying, will enter into the

set of the assigned conditions α_1, α_2, ...

For example, during the optimization of the job schedule of industrial enterprise it is possible to require so that the gain would be maximum, plan/layout with respect to assortment was fulfilled, but the prime cost of production - not higher than given one. During gliding/planning of the bombing raid, it is possible to require so that the replaced to the enemy damage would be maximum, but in this case, their own losses and the cost/value of operation did not exceed known limits.

Page 27.

Upon this formulation of the problem, all indices of efficiency, except one, main, are translated into the discharge of the assigned conditions of operation. The versions of solution, which are not placed in the assigned boundaries, immediately are reject/thrown as noncompetitive. The obtained recommendations, obviously, will depend on that, as are selected limitations for auxiliary indices. In order to determine, how this affects the final recommendations by choice of solution, it is useful to vary limitation within reasonable limits.

Finally, is feasible one additional way of the construction of compromise solution, which can be named the "method of consecutive concessions".

Let us assume that the indices of efficiency are arrange/located in the order of the decreasing importance: first basic W_1, then others, auxiliary: W_2, W_3, ... For simplicity let us consider that each of them must be converted into the maximum (if this not then, it suffices to change the sign of index). The procedure of the construction of compromise solution is reduced to following. First is ought the solution, which rotates into maximum the main index of efficiency W_1. Then it is assigned, on the basis of practical considerations and accuracy, from which are known initial data (but frequently it is small), certain "concession" ΔW_1 which we are concordant to allow in order to convert into maximum the second index W_2. We assign on index W_1 limitation so that it would be not less than $W^*_1 - \Delta W_1$, where W^*_1 - maximally possible value W_1, and during this limitation we seek the solution, which rotates into maximum W_2. Further again is assigned the "concession" in index W_2, by value of which it is possible to maximize W_3, and so forth.

This method of the construction of compromise solution is good in that it is here immediately evident, by the value of which "concession" in one index is acquired gain in other.

Let us note that the freedom of choice of solution, acquired by

the value even of insignificant "concessions", can render/show essential, since in maximum region usually the efficiency of solution varies very weakly.

One way or another, with any method of formalization, the problem of the quantitative proof of solution by several indices remains not to end determined, and the final selection of solution is determined "commander's volitional event/report" (so we conditionally will call the critical for selection face). Researcher's matter - to let into commander's order a sufficient quantity of data, that permit for it to thoroughly consider advantages and deficiency/lacks in each version of solution and, relying on them, to do a final selection.

Page 28.

2. LINEAR PROGRAMMING.

1. Problems of linear programming.

In many regions of practice, appear the peculiar problems of optimization of the solutions, for which are characteristic the following features:

- index of efficiency W represents by itself linear function from the cell/elements of solution $x_1, x_2, \ldots$;

- limiting conditions, assigned for the possible solutions, take the form of linear equalities or inequalities.

Such problems is conventionally designated as the problems of linear programming [1].

FOOTNOTE [1]. Word "programming" borrowed from foreign literature and in this case indicates nothing else but "gliding/planning". ENDFOOTNOTE.

Let us give several examples of the problems of linear programming of the different regions of practice.

1. Problem of food ration. There are four forms of food products:

$$\Pi_1, \Pi_2, \Pi_3, \Pi_4.$$

Is known the cost/value of unity of each product:

Of these products it is necessary to comprise food ration which must contain:

$$c_1, c_2, c_3, c_4.$$

- proteins is not less than b_1 unity,

- carbohydrates are not less than b_2 unity,

- grease are not less than b_3 unity.

Unity of product Π_1 contains a_{11} unity of proteins, a_{12} unity of carbohydrates, a_{13} unity of grease, etc. The content of cell/elements in unity of each product is assigned by table (Table 1.1).

It is required so to comprise food ration in order to ensure assigned conditions (1.1) with the minimum cost/value of ration.

Let us write the formulated verbal conditions of problem in the form of mathematical formulas. Let us designate

$$x_1, x_2, x_3, x_4$$

quantities of products $\Pi_1, \Pi_2, \Pi_3, \Pi_4,$ entering the ration.

Page 29.

It is obvious, the common/general/total cost/value of the ration will be

$$L = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$$

or it is shorter

$$L = \sum_{i=1}^{4} c_i x_i. \qquad (1.2)$$

Let us register mathematically conditions (1.1). In one unity of product Π_1 is contained a_{11} unity of protein, which means, that in x_1 unity - $a_{11}x_1$; in x_2 unity of product Π_2 is contained $a_{21}x_2$ unity of protein, etc. The total quantity of proteins, which is contained in ration, must not be less than b_1; we hence obtain the first condition-inequality:

$$a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4 \geqslant b_1. \qquad (1.3)$$

Record/writing analogous conditions for carbohydrates and fats, we will obtain, including (1.3), three condition-inequality:

$$\left.\begin{array}{l} a_{11}x_1+a_{21}x_2+a_{31}x_3+a_{41}x_4 \geqslant b_1, \\ a_{12}x_1+a_{22}x_2+a_{32}x_3+a_{42}x_4 > b_2. \\ a_{13}x_1+a_{23}x_2+a_{33}x_3+a_{43}x_4 \geqslant b_3. \end{array}\right\} \quad (1.4)$$

These conditions represent by themselves the limitations, superimposed on the solution.

Appears the following task:

to select such nonnegative values of variables $x_1, x_2, x_3, x_4,$ that satisfy linear inequalities (1.4), with which the linear function of these variables

$$L = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$$

would be converted into minimum.

Stated problem represents by itself the typical task of linear programming. Without being stopped as far as on the methods of its solution (about this speech it will go subsequently), let us place still several similar tasks.

Table 1.1.

		(1) Элемент		
		(2) белки	(3) углеводы	(4) жиры
(5) Продукт	$П_1$	a_{11}	a_{12}	a_{13}
	$П_2$	a_{21}	a_{22}	a_{23}
	$П_3$	a_{31}	a_{32}	a_{33}
	$П_4$	a_{41}	a_{42}	a_{43}

Key: (1). Cell/element. (2). proteins. (3). carbohydrates. (4). grease. (5). Product.

Page 30.

2. Task of loading of machine tools. Weaving factory has available N_1 machine tools of type 1 and N_2 machine tools of type 2. Machine tools can produce four forms of the fabrics:

$$T_1, T_2, T_3, T_4.$$

Each type of machine tool can produce any of the forms of fabrics, but in unidentical quantity. A machine tool of type 1 is produced in month a_{11} of the meters of fabric T_1, a_{12} meters of fabric T_2, a_{13} meters of fabric T_3, a_{14} meters of fabric T_4. The corresponding numbers for a machine tool of type 2 will be a_{21}, a_{22}, a_{23}, a_{24}. Thus, the productivity of machine tools in the production of each form of fabric are assigned to Table 1.2.

Each meter of fabric T_1 yields to factory income c_1, fabrics T_2

- income c_2, fabrics T_3 - income c_3 and fabrics T_4 - income c_4. To factory is prescribed, the plan/layout according to which it is due to produce for the month:

is not less than b_1 meters of fabric T_1, is not less than b_2 meters of fabric T_2, is not less than b_3 meters of fabric T_3 and is not less than b_4 meters of fabric T_4, i.e., planned target is expressed by numbers b_1, b_2, b_3, b_4.

It is required so to distribute loading machine tools by the production of the fabrics of different form so that the plan/layout would be carried out and in this case monthly gain would be maximum.

Let us register the conditions of task mathematically. Let us designate x_{11} - number of machine tools of type 1, occupied with the production of fabric T_1, x_{12} - number of machine tools of type 1, occupied with the production of fabric T_2, and generally x_{ij} — a number of machine tools of the type i, occupied with the production of fabric T_j. The first index corresponds to the type of machine tool, the second - to a form of fabric ($i = 1, 2$, $j = 1, 2, 3, 4$).

Thus appear eight variables - cell/elements of the solution:

$$\left.\begin{array}{l} x_{11},\ x_{12},\ x_{13},\ x_{14}; \\ x_{21},\ x_{22},\ x_{23},\ x_{24}, \end{array}\right\} \qquad (1.5)$$

which we they must select, so that the monthly gain would be maximum.

Let us register formula for calculating this gain. Each meter of fabric T_1 yields gain c_1; x_{11} meters of fabric T_1 they will bring gain c_1x_{11}; in all fabric ^{1} will bring gains $c_1(x_{11} + x_{21})$ and so forth. Common/general/total gain will be equal to:

$$L = c_1(x_{11} + x_{21}) + c_2(x_{12} + x_{22}) + c_3(x_{13} + x_{23}) + c_4(x_{14} + x_{24}). \quad (1.6)$$

It is required to select such nonnegative values of variables (1.5), so that the linear function from them (1.6) would be converted into maximum.

Table 1.2.

(1) Тип станка	(2) Вид ткани			
	T_1	T_2	T_3	T_4
1	a_{11}	a_{12}	a_{13}	a_{14}
2	a_{21}	a_{22}	a_{23}	a_{24}

Key: (1). Type of machine tool. (2). Form of fabric.

Page 31.

In this case, must be implemented the following limiting conditions:

1) Resource/lifetimes on machine tools must not be exceeded, i.e., the sum of quantities of each type machine tools, occupied with the production of all fabrics, must not exceed the available stock of the machine tools:

$$\left.\begin{aligned} x_{11}+x_{12}+x_{13}+x_{14} &\leqslant N_1; \\ x_{21}+x_{22}+x_{23}+x_{24} &\leqslant N_2. \end{aligned}\right\} \quad (1.7)$$

2) Assignments with respect to assortment must be carried out (or are exceeded). With account of the data of Table. 1.2 these conditions will be registered in the form of the inequalities:

$$\left.\begin{aligned} a_{11}x_{11}+a_{21}x_{21} &\geqslant b_1, \\ a_{12}x_{12}+a_{22}x_{22} &\geqslant b_2, \\ a_{13}x_{13}+a_{23}x_{23} &\geqslant b_3, \\ a_{14}x_{14}+a_{24}x_{24} &\geqslant b_4. \end{aligned}\right\} \quad (1.8)$$

Thus, is formulated the task:

To select such non-negative values of variables x_{11}, x_{12}, ..., ..., x_{24}, that satisfy linear inequalities (1.7) and (1.8), by which

the linear function of these variables (1.6) would be converted into maximum.

3. Task of distribution of resource/lifetimes. There are some resource/lifetimes (raw material, work force, equipment):

$$R_1, R_2, \ldots, R_m$$

in quantities respectively

$$b_1, b_2, \ldots, b_m$$

units. With the help of these resource/lifetimes can be produced the goods:

$$T_1, T_2, \ldots, T_n.$$

For the production of one unity of goods T_j it is necessary a_{ij} to unity of resource/lifetime R_i (i = 1, 2, ..., m; j = 1, 2, ..., n). Each unity of resource/lifetime R_i costs d_i rubles (i = 1, 2, ..., m). Each unity of goods T_j can be realized on value $c_j (j = 1, 2, \ldots, n)$.

In each form of goods a quantity of produced unity is limited to the demand: it is known that the market cannot absorb more than k_j unity of goods T_j $(j = 1, 2, \ldots, n)$.

It does ask itself: which quantity of unity of which goods must be produced, in order to realize maximum gain?

Let us register the conditions of task. Let us designate

$$x_1, x_2, \ldots x_n$$

quantities of goods T_1, T_2, ..., T_n, which we will plan to production. The conditions of demand assign on these values of the limitation:

$$x_1 \leqslant k_1; \quad x_2 \leqslant k_2; \ldots; \quad x_n \leqslant k_n. \qquad (1.9)$$

Page 32.

Resource/lifetimes must suffice, hence appear the limitations:

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leqslant b_1; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leqslant b_2; \\ \cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leqslant b_m. \end{array}\right\}$$

These conditions it is possible to write more briefly in the form:

$$\left.\begin{array}{l} \sum_{j=1}^{n} a_{1j}x_j \leqslant b_1, \\ \sum_{j=1}^{n} a_{2j}x_j \leqslant b_2, \\ \cdots\cdots \\ \sum_{j=1}^{n} a_{mj}x_j \leqslant b_m. \end{array}\right\} \qquad (1.10)$$

It is expressed gain L depending on the cell/elements of the solution

$$x_1, x_2, \ldots, x_n.$$

Prime cost s_j of unity of goods T_j is equal to

$$s_j = a_{1j}d_1 + a_{2j}d_2 + \ldots + a_{mj}d_m,$$

or, it is shorter,

$$s_j = \sum_{i=1}^{m} a_{ij}d_i \quad (j = 1, 2, \ldots, n) \qquad (1.11)$$

After computing according to this formula the prime cost of

unity of each goods, we will obtain a series of the values:

$$s_1, s_2, \ldots, s_n.$$

Pure/clean gain q_j, obtained from the realization of one unity of goods T_j, is equal to the difference between its selling-price c_j and prime cost s_j:

$$q_j = c_j - s_j \quad (j = 1, 2, \ldots, n). \qquad (1.12)$$

On this formula we obtain pure/clean gains per unit for all goods:

$$q_1, q_2, \ldots, q_n.$$

Common/general/total pure/clean gain from the realization of all goods will be

$$L = q_1 x_1 + q_2 x_2 + \ldots + q_n x_n,$$

or, it is shorter,

$$L = \sum_{j=1}^{n} q_j x_j. \qquad (1.13)$$

Page 33.

Task is reduced to following:

To select such nonnegative values of the variables $x_1, x_2, \ldots, x_n$ which satisfy linear inequalities (1.9), (1.10) and convert into maximum the linear function of these variables (1.13).

4. Transport problem. There are m of the storages:

$$C_1, C_2, \ldots, C_m$$

and n of the point/items of the consumption:

$$\Pi_1, \Pi_2, \ldots, \Pi_n$$

(see Fig. 2.1).

Speech occurs about the composition of the plan/layout of transport from storages $C_1, C_2, \ldots, C_m$ into point/items $\Pi_1, \Pi_2, \ldots, \Pi_n$ certain goods. On storages $C_1, C_2, \ldots, C_m$ are supplies of goods in the quantities

$$a_1, a_2, \ldots, a_m$$

of unity. The point/items of consumption $\Pi_1, \Pi_2, \ldots, \Pi_n$ fed the claims respectively to

$$b_1, b_2, \ldots, b_n$$

of unity of goods. Claims are feasible, i.e., the sum of all claims does not exceed the sum of all available supplies:

$$\sum_{j=1}^{n} b_j \leqslant \sum_{i=1}^{m} a_i.$$

Storages $C_1, \ldots, C_m$ are connected with the point/items of consumption $\Pi_1, \ldots, \Pi_n$ by some grid/network of roads with the specific tariffs on transport. The cost/value of the transport of one unity of goods from storage C_i into point/item Π_j is equal to c_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$).

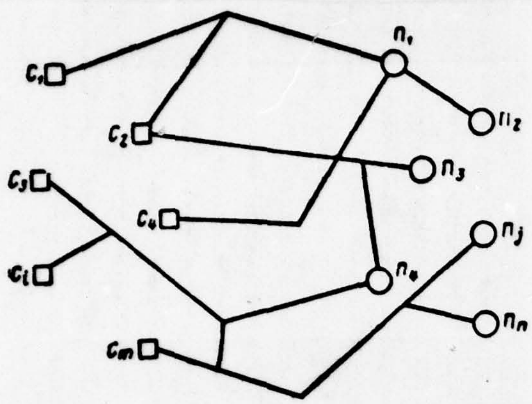


Fig. 2.1.

Page 34.

It is required to comprise the plan/layout of transport, i.e., to indicate that from which storage into which point/items and which quantity of goods must be directed so that the claims would be carried out, and overall expenditure/consumptions to all transport were minimum.

Let us designate x_{ij} — the quantity of unity of goods, directed from storage C_i in point/item Π_j (if from this storage for this point/item goods are not directed, $x_{ij}=0$).

The solution (plan/laycut of transport) consists of Mn of the numbers:

$$\begin{array}{l} x_{11}, x_{12}, \dots, x_{1n}; \\ x_{21}, x_{22}, \dots, x_{2n}; \\ \dots\dots\dots\dots \\ x_{m1}, x_{m2}, \dots, x_{mn}, \end{array}$$

forming rectangular array (matrix/die). Let us in abbreviated form designate it (x_{ij}). It is required to select such nonnegative values of variables x_{ij} (i = 1, 2, ..., m; j = 1, 2, ..., n) so that would be satisfied the following conditions:

1. The capacitance/capacity of storages must not be exceeded, i.e., the total quantity of goods, undertaken from each storage, must not exceed the available on it supplies:

$$\left.\begin{array}{l} x_{11}+x_{12}+\ldots+x_{1n}\leqslant a_1; \\ x_{21}+x_{22}+\ldots+x_{2n}\leqslant a_2; \\ \cdots\cdots\cdots\cdots \\ x_{m1}+x_{m2}+\ldots+x_{mn}\leqslant a_m, \end{array}\right\}$$

or, it is shorter,

$$\left.\begin{array}{l} \sum_{j=1}^{n} x_{1j}\leqslant a_1, \\ \sum_{j=1}^{n} x_{2j}\leqslant a_2, \\ \cdots\cdots \\ \sum_{j=1}^{m} x_{mj}< a_m. \end{array}\right\} \quad (1.14)$$

2. Claims, subject by point/items of consumption, must be carried out:

$$\left.\begin{array}{l} x_{11}+x_{21}+\ldots+x_{m1}=b_1, \\ x_{12}+x_{22}+\ldots+x_{m2}=b_2, \\ \cdots\cdots\cdots\cdots \\ x_{1n}+x_{2n}+\ldots+x_{mn}=b_n \end{array}\right\}$$

either it is shorter or, it is shorter,

$$\left.\begin{array}{l} \sum_{i=1}^{m} x_{i1}=b_1, \\ \sum_{i=1}^{m} x_{i2}=b_2, \\ \cdots\cdots \\ \sum_{i=1}^{m} x_{in}=b_n. \end{array}\right\} \quad (1.15)$$

Page 35.

The common/general/total cost/value of transport L will be equal to

$$L = c_{11} x_{11} + c_{12} x_{12} + \dots + c_{1n} x_{1n} + \\ + c_{21} x_{21} + c_{22} x_{22} + \dots + c_{2n} x_{2n} + \\ + \dots + c_{m1} x_{m1} + c_{m2} x_{m2} + \dots + c_{mn} x_{mn},$$

or, it is much shorter,

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}. \qquad (1.16)$$

It is required so to select the plan/layout of transport (x_{ij}) (i = 1, 2, ..., m; j = 1, 2, ..., n) in order cost/value L of these transport to convert into the minimum.

Again arises the problem, analogous examined earlier: to select non-negative values of variables (x_{ij}) so that during satisfaction of conditions (1.14), (1.15) the linear function of these variables (1.16) would reach the minimum.

Certain special feature/peculiarity of this problem, in comparison with those previously examined, lies in the fact that not

all restrictions placed on variables are linear inequalities; namely, conditions (1.15) are registered in the form of linear equalities.

In the future we will meet the problems of the linear programming in which limiting conditions take both the form of linear inequalities and equalities, and will learn with lightness/ease to pass from some to others and vice versa.

Let us note that upon certain setting of transport problem all limiting conditions of problem become equalities. Namely, if the sum of all claims equal to the sum of all supplies

$$\sum_{j=1}^{n} b_j = \sum_{i=1}^{m} a_i,$$

then each storage will be unavoidably from exported everything which on it is, and inequality (1.14), just as (1.15), after being converted into equalities.

This problem about transport is called transport problem, and it we will be specially occupied subsequently (see §9-14 this chapters).

5. Problem of production of complex equipment. Plan/glides the production of the complex equipment whose each assembly consists of n of the cell/elements:

$$Э_1, Э_2, \dots, Э_n.$$

Page 36.

Orders to the production of these cell/elements can be placed on ■ the different enterprises:

$$\Pi_1, \Pi_2, \ldots, \Pi_m.$$

During preset time T in enterprise Π_i it is possible to prepare a_{ij} cell/elements of type $Э_j$ (i = 1, ..., m; j = 1, ..., n).

To delivery are subject only full/total/complete assemblies of equipment, which consist of the set of all cell/elements $Э_1, Э_2, \ldots, Э_n$.

It is required to distribute orders on enterprises so that the number of full/total/complete assemblies of equipment, prepared for time T, would be maximal. Plan/gliding the production of equipment, we must for each enterprise Π_i indicate, what part of the available in its order time it must return to the production of cell/elements $Э_j$ (i = 1, ..., m; j = 1, ..., n).

Let us designate x_{ij} the fraction of time T which enterprise Π_i it will give to the production of cell/element $Э_j$ (if this cell/element in this enterprise not at all is produced, $x_{ij}=0$).

During gliding/planning we must observe the following limiting

conditions: the quantity of time which each enterprise spends on the production of all cell/elements, must not exceed the overall supply of time T (but "fraction" - unity):

$$\left.\begin{aligned} x_{11}+x_{12}+\ldots+x_{1n}\leqslant 1,\\ x_{21}+x_{22}+\ldots+x_{2n}\leqslant 1,\\ \ldots\ldots\ldots\ldots\\ x_{m1}+x_{m2}+\ldots+x_{mn}\leqslant 1,\end{aligned}\right\}$$

or

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{1j}\leqslant 1,\\ \sum_{j=1}^{n} x_{2j}\leqslant 1,\\ \ldots\ldots\\ \sum_{j=1}^{n} x_{mj}\leqslant 1.\end{aligned}\right\} \qquad (1.17)$$

Let us determine a quantity of full/total/complete assemblies of the equipment which for time T will place all enterprises together.

The total quantity of cell/elements which will produce all $Э_j$, enterprises together, will be equal

$$N_j = a_{1j}x_{1j} + a_{2j}x_{2j} + \ldots + a_{mj}x_{mj}$$

or

$$N_j = \sum_{i=1}^{m} a_{ij}x_{ij} \quad (j = 1, \ldots, n). \qquad (1.18)$$

Page 37.

Thus, with the assigned plan/layout of the distribution of orders, i.e., with given ones x_{ij} (i = 1, ..., m; j = 1, ..., n) will be produced:

—N_1 (1) экземпляров элемента $Э_1$

—N_2 (1) экземпляров элемента $Э_2$

. . . .

—N_n (1) экземпляров элемента $Э_n$.

Key: (1). The copies of cell/element.

How many complete assemblies of equipment it is possible to gather from these elements? Is obvious so many, how minimum of all numbers N_1, N_2, ..., N_n. It is real/actual, if, for example, elements of the type $_1$ is produced by 100 pcs., and cell/elements of the type $Э_2$ — a total of of 10 pcs., then we in any way can gather of these cell/elements of more than 10 full/total/complete assemblies.

Let us designate Z - quantity of full/total/complete assemblies of the equipment which can be gathered with this plan/layout of the arrangement/position of orders (x_{ij}).

We have:

$$Z = \min_j N_j, \qquad (1.19)$$

where by sign $\min_j$ is designated minimum from the numbers, which stand under this sign, for all possible j.

Taking into account (1.18), condition (1.19) can be rewritten in the form

$$Z = \min_j \sum_{i=1}^{m} a_{ij} x_{ij}. \qquad (1.20)$$

Thus, we come to the following mathematical formulation of the problem:

To find such nonnegative values of variables x_{ij}, so that would be implemented inequalities (1.17) and in this case was converted into maximum the function of these variables

$$Z = \min_{j} \sum_{i=1}^{m} a_{ij} x_{ij}.$$

The difference for this problem from all those previously examined lies in the fact that here maximized function Z is not linear function from variables x_{ij} and, thus, problem, strictly, is not the problem of linear programming. However, it it is easy to reduce to the problem of linear programming by following reasonings.

Since value Z is minimum of all values $N_j = \sum_{i=1}^{m} a_{ij} x_{ij}$, then it is possible to write a series of the inequalities

$$\left.\begin{array}{l} \sum_{i=1}^{m} a_{i1} x_{i1} \geqslant Z; \\ \sum_{i=1}^{m} a_{i2} x_{i2} \geqslant Z; \\ \cdots\cdots \\ \sum_{i=1}^{m} a_{in} x_{in} \geqslant Z. \end{array}\right\} \quad (1.21)$$

Page 38.

Value Z can be considered as new nonnegative variable and to solve following problem.

To find such non-negative values of the variables x_{11}, x_{12}, ..., x_{mn}

To find such nonnegative values of variables x_{ij}, so that would be implemented inequalities (1.17) and in this case was converted into maximum the function of these variables

$$Z = \min_{j} \sum_{i=1}^{m} a_{ij} x_{ij}.$$

The difference for this problem from all those previously examined lies in the fact that here maximized function Z is not linear function from variables x_{ij} and, thus, problem, strictly, is not the problem of linear programming. However, it it is easy to reduce to the problem of linear programming by following reasonings.

Since value Z is minimum of all values $N_j = \sum_{i=1}^{m} a_{ij} x_{ij}$, then it is possible to write a series of the inequalities

$$\left.\begin{array}{l} \sum_{i=1}^{m} a_{i1} x_{i1} \geqslant Z; \\ \sum_{i=1}^{m} a_{i2} x_{i2} \geqslant Z; \\ \cdots\cdots \\ \sum_{i=1}^{m} a_{in} x_{in} \geqslant Z. \end{array}\right\} \quad (1.21)$$

Page 38.

Value Z can be considered as new nonnegative variable and to solve following problem.

To find such non-negative values of the variables x_{11}, x_{12}, ..., x_{mn}

and Z, so that they would satisfy linear inequalities (1.17) and (1.21) and with this value Z it was converted into maximum.

Since value Z is a linear function of the new variables x_{11}, x_{12}, ..., x_{mn}, Z:

$$Z=0\cdot x_{11}+0\cdot x_{12}+\ldots+0\cdot x_{mn}+1\cdot Z,$$

the problem is reduced to the usual problem of linear programming, by introduction "excess" alternating/variable Z which in the initial formulation of the problem did not figure.

Problems of such type where it is required to convert into maximum the minimum value of some value (or, on the contrary, into minimum-maximum), fairly often they are encountered in practice and are called "problems to minimax". With such problems we will be still met in Chapter 10.

Thus, we considered a whole series of the problems of operations research from the different regions of practice; these problems are characterized by some common/general/total features. In each of them the cell/elements of solution represent by themselves a series of the nonnegative variables x_1, x_2, ... Is required so to select the values of these variables, so that

1) would be implemented some limitations, having the form of

linear inequalities or equalities relative to the variables x_1, x_2, ...;

2) certain linear function L the same variables it was converted into maximum (minimum).

The mathematical apparatus of the linear programming presentation of which we begin, is intended specially for the solution of such problems.

Can arise the question: a is necessary this special apparatus? It is cannot whether, as is customary in mathematics, it is simple to differentiate L arguments x_1, x_2, ..., to make derived equal to zero and to solve the obtained system of equations?

No, it turns out that it is not possible to do this ! Since function L is linear, derivatives of it on all arguments are constant and nowhere into zero they are converted. The maximum (or the minimum) of function L, if it exists, it is reached always somewhere on the boundary of the region of the possible values x_1, x_2, ..., i.e., where come into action limitations.

Page 39.

The mathematical apparatus of linear programming allows for us consecutively, within the shortest periods, to examine the boundaries of the region of possible sclutions and to find on these boundaries the solution, which is optimum, i.e., such value part x_1, x_2, ..., with which the linear function L is converted into maximum or into the minimum.

2. Basic problem of linear programming.

Above we considered different practical prcblems, which were being reduced to the diagram of linear programming. In some of these problems linear limitaticns took the form of inequality, in others - equalities, in the third - those and others.

Here we will examine the problem cf linear programming with limitation-equalities - the so-called basic prcblem of linear programming (OZLP).

In the future we will show how frcm prcblem with limitation-inequalities it is possible to pass to OZLP, and vice versa.

The basic problem of linear programming is placed as follows.

There is a series of variables

$$x_1, x_2, \dots, x_n.$$

It is required to find such nonnegative values of these variables which would satisfy the system of the linear equations:

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1; \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2; \\ \dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \end{aligned}\right\} \quad (2.1)$$

and, furthermore, would be converted into minimum the linear function

$$L = c_1x_1 + c_2x_2 + \dots + c_nx_n. \quad (2.2)$$

It is obvious, the case when linear function must be converted not into the minimum, but into maximum, easily it is reduced to previous, if we change the sign of function and to consider instead of it the function

$$L' = -L = -c_1x_1 - c_2x_2 - \dots - c_nx_n. \quad (2.3)$$

Let us agree to call the permissible solution of OZLP any aggregate of variables

$$x_1 \geqslant 0, \ x_2 \geqslant 0, \dots, \ x_n \geqslant 0,$$

satisfying equations (2.1).

Optimum solution let us call that of the permissible solutions, with which linear function (2.2) is converted into the minimum.

Page 40.

The basic problem of linear programming not necessarily must have solution. It can seem that equations (2.1) contradict each other; it can seem that they have solution, but not in the range of the nonnegative values $x_1, x_2, \ldots, x_n$. Then OZLP does not have the permissible solutions. Finally, it can seem that permissible solutions of OZLP exist, but among them no optimum: function L in the domain of the permissible solutions is not limited from below.

With examples of such features of OZLP we will be introduced subsequently.

Let us consider, first of all, a question concerning the existence of the permissible solutions of OZLP.

During the solution of this question, we can exclude from examination the linear function L which is required to minimize - the presence of the permissible solutions it is determined only by equations (2.1).

Thus, let there be system of equations (2.1). Are there the non-negative values $x_1, x_2, \ldots, x_n$, satisfying this system? This question is examined in the special section of mathematics - linear algebra.

Let us give briefly some positions of linear algebra, not stopping on the proofs of the corresponding theorems [1].

FOOTNOTE [1]. The elementary presentation of linear algebra see, for example, in work [5]. ENDFOOTNOTE.

The matrix/die of the system of the linear equations

$$\left.\begin{array}{l} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n=b_1;\\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n=b_2;\\ \cdots\cdots\cdots\cdots\cdots\\ a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n=b_m \end{array}\right\}$$

is called the table, comprised of the coefficients of $x_1, x_2, \ldots, x_n$:

$$\begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \cdot & \cdot & \cdot & \cdot\\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{Vmatrix}$$

The augmented matrix of the system of linear equations is called the same matrix/die, supplemented by the column of the absolute terms:

$$\begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1\\ a_{21} & a_{22} & \ldots & a_{2n} & b_2\\ \cdot & \cdot & \cdot & \cdot & \cdot\\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{Vmatrix}$$

The rank of matrix/die is called the greatest order of different from zero definitions which can be obtained, deleting from matrix/die some rows and some columns.

In linear algebra is proven, which for the consistency of the system of linear equations (2.1) is necessary and it is sufficient so that the rank of the matrix/die of system would be equal to the rank

of its augmented matrix.

Page 41.

Example 1. Is given the system of three equations with four unknowns:

$$\left.\begin{aligned} 2x_1+x_2-x_3+x_4&=-1;\\ x_1-x_2&=2;\\ x_1-2x_3&=3.\end{aligned}\right\}$$

To determine, is this system of combined?

Solution. Matrix/die of the system:

$$\left\|\begin{matrix} 2 & 1 & -1 & 1\\ 1 & -1 & 0 & 0\\ 1 & 0 & -2 & 0\end{matrix}\right\|.$$

The augmented matrix:

$$\left\|\begin{matrix} 2 & 1 & -1 & 1 & -1\\ 1 & -1 & 0 & 0 & 2\\ 1 & 0 & -2 & 0 & 3\end{matrix}\right\|.$$

Let us determine the rank of the first matrix/die. It cannot be more than 3 (since the number of rows is equal to 3). Let us comprise any definition, eliminating from the matrix/die any column, for example, the latter. We will obtain

$$\Delta=\begin{vmatrix} 2 & 1 & -1\\ 1 & -1 & 0\\ 1 & 0 & -2\end{vmatrix}.$$

Computing this determinant according to known rule, we will obtain:

$$\Delta=2\cdot(-1)\cdot(-2)+1\cdot0\cdot(-1)+1\cdot1\cdot0-\\-(-1)\cdot(-1)\cdot1-2\cdot0\cdot0-(-2)\cdot1\cdot1=4-1+2=5.$$

This determinant is not equal to zero, which means, that the rank of the matrix/die of system is equal to 3. Obviously, the rank of augmented matrix is also equal to 3, since of the cell/elements of augmented matrix it is possible to comprise the same determinant. From equality the ranks of matrix/dies, it follows that the system of equations is combined.

Example 2. To trace to consistency the system of two equations with three unknowns x_1, x_2, x_3:

$$2x_1 - x_2 + x_3 = -4,$$
$$4x_1 - 2x_2 + 2x_3 = 1.$$

Solution. The augmented matrix of the system:

$$\left\| \begin{matrix} 2 & -1 & 1 & -4 \\ 4 & -2 & 2 & 1 \end{matrix} \right\|.$$

(her left side - matrix/die of system).

Let us find the rank of the matrix/die of system, comprising all the possible determinants of the second order:

$$\Delta_1 = \begin{vmatrix} 2 & -1 \\ 4 & -2 \end{vmatrix} = -4+4=0;$$

$$\Delta_2 = \begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix} = 4-4=0;$$

Page 42.

$$\Delta_3 = \begin{vmatrix} -1 & 1 \\ -2 & 2 \end{vmatrix} = -2+2=0.$$

Thus all the possible determinants of the second order, the comprised of matrix elements of system are equal to zero; that means the rank of this matrix/die of system $r_c=1<2$.

Let us find the rank of augmented matrix. Determinant

$$\Delta_4 = \begin{vmatrix} 2 & -4 \\ 4 & 1 \end{vmatrix} = 2 + 16 = 18 \neq 0.$$

Hence the rank of augmented matrix $r_p = 2$, it is not equal to the rank of the matrix/die of system; $r_p \neq r_c$. Consequently, system of equations is incompatible/inconsistent.

Example 3. To trace to consistency the system of three equations with four unknowns:

$$\begin{aligned} x_1 + x_2 + x_3 - x_4 &= 2, \\ x_1 - x_2 + x_3 + x_4 &= -1, \\ 3x_1 - x_2 + 3x_3 + x_4 &= 0 \end{aligned}$$

Solution. The augmented matrix of system (together with the matrix/die of the system):

$$\left\| \begin{array}{rrrrr} 1 & 1 & 1 & -1 & 2 \\ 1 & -1 & 1 & 1 & -1 \\ 3 & -1 & 3 & 1 & 0 \end{array} \right\|.$$

Let us find the rank of the matrix/die of system. Let us take the determinant of the third order, comprised of its cell/elements, for example:

$$\Delta_1 = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 3 & -1 & 3 \end{vmatrix}.$$

It is known that if any row of the determinant is the linear combination other of its other rows, then determinant is equal to zero. In our case the third row is linear combination the first two: in order it to obtain that is sufficient to sum the first row from that doubled the second Therefore $\Delta_1 = 0$.

Is not difficult to ascertain that the same property it possesses and any determinant of the third order, comprised of the matrix elements of system. Consequently, the rank of matrix/die system $r_c < 3$.

Since there is nonzero determinant of the second order, for example,

$$\Delta_2 = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2,$$

that the rank of the matrix/die of system is equal to $r_c = 2$.

With the help of the same reasonings let us ascertain that and the rank of augmented matrix is equal to twos; $r_p = 2$. Consequently, system of equations is combined.

Let us note that three equations of this example are not independent variables: the third can be obtained of the first two, if we multiply the second by two and to adjoin to the first. That means that the third equation is a simple corollary the first two. Independent equations in system only dv: this is also reflected by the fact that the rank of the matrix/die of system $r_c = 2$.

Thus, if the system of equation-limitations of OZLP is combined, then the matrix/die of system and its augmented matrix have one and the same rank.

Page 43.

This common/general/total rank r is called the rank of system; it represents by itself nothing else but the number of linearly independent equations among the superimposed limitations.

It is obvious, the rank of system r cannot be more than the number of equations m:

$$r \leqslant m.$$

It is obvious, also that the rank of system cannot be more than the total number of variables n:

$$r \leqslant n.$$

It is real/actual, the rank of the matrix/die of system is defined as greatest order of the determinant, comprised of matrix elements; since the number of its rows is equal to m, the $r \leqslant m$; since the number of its columns is equal to n, the $r \leqslant n$.

The structure of the problem of linear programming depends substantially on the rank of the system of limitations (2.1).

Let us consider, first of all, the case when r = n, i.e., when the number of the linearly independent equations, entering system

(2.1), equal to number of variables n. Let us reject/throw the "excess" equations, which are the linear combinations of others. The system of equation-limitations of OZLP takes the form:

$$\left.\begin{array}{l} a_{11}x_1+a_{12}x_2+\ldots+a_{1i}x_i+\ldots+a_{1n}x_n=b_1, \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2i}x_i+\ldots+a_{2n}x_n=b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1+a_{n2}x_2+\ldots+a_{ni}x_i+\ldots+a_{nn}x_n=b_n. \end{array}\right\} \quad (2.4)$$

Since r = n, then the definition, compiled from coefficients,

$$\Delta=\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1i} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2i} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \ldots & a_{ni} & \ldots & a_{nn} \end{vmatrix}$$

is not equal to zero. From algebra it is known that in this case system (2.4) has unique solution. In order to find value x_i, it suffices in determinant Δ to replace the i column - by column of absolute terms and to divide into Δ.

Thus, with r = n the system of equation-limitations of OZLP has only the solution:

$$x_1, x_2, \ldots, x_n.$$

If in this solution at least one of values $x_1, x_2, \ldots, x_n$ is negative, this means that the obtained solution is inadmissible and, which means, that OZLP does not have solution.

If all values $x_1, x_2, \ldots, x_n$ are nonnegative, then the obtained solution is permissible. It is obvious it is and optimum (because there are no others).

Page 44.

It is obvious, this trivial case cannot us interest. Therefore subsequently we will examine only case when $r < n$, i.e., when the number of independent equations, by which they must satisfy the variables x_1, x_2, ..., x_n, lesser than the number of variables themselves. Then, if system is combined, of it there is a countless solution set. With this $n - r$ alternating/variable we can assign arbitrary values (the so-called unrestricted variables), and the others r of variables will be expressed by them (these r of variables we will call base).

Example 4. Is examined the system of two equations with four unknowns:

$$\left.\begin{aligned} 2x_1 - x_2 + x_3 - x_4 &= 1, \\ -x_1 + x_2 - 2x_3 + x_4 &= 2. \end{aligned}\right\} \quad (2.5)$$

The rank of this system is equal $r = 2$ (equations are linearly independent). Let us select as unrestricted variables, e.g., x_3 and x_4, and as basic - x_1 and x_2. Let us express the basic variables in terms of the unrestricted ones. We have from equations (2.5):

$$\begin{aligned} 2x_1 - x_2 &= 1 - x_3 + x_4, \\ -x_1 + x_2 &= 2 + 2x_3 - x_4 \end{aligned}$$

Store/adding up these equations, we will obtain

$$x_1 = 3 + x_3.$$

Multiplying the second equation to 2 and store/adding up with the first, we will obtain

$$x_2 = 5 + 3x_3 - x_4.$$

AD-A066 460 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/2
OPERATIONS RESEARCH. PART I,(U)
JUN 78 Y S VENTTSEL'
UNCLASSIFIED FTD-ID(RS)T-0687-78-PT-1 NL

2 OF 6
AD
A066460

Thus, the base variables x_1, x_2 are expressed through the free x_3, x_4. By unrestricted variable x_3, x_4 it is possible to give any values; in this case, we will always obtain the value part x_1, x_2, x_3, x_4, which satisfies system of equations (2.5). For example, set/assuming $x_3 = x_4 = 0$, we will obtain $x_1 = 3$, $x_2 = 5$; these values satisfy system (2.5). Set/assuming $x_3 = 1$, $x_4 = 2$, we will obtain $x_1 = 4$, $x_2 = 6$; these values also satisfy equations (2.5).

Generally, if the rank of system of equations of OZLP (i.e. the number of the linearly independent equations, entering limitation system) is equal r, then always it is possible to express some r of the base alternating/variable through n - r others (free) and, giving to unrestricted variable any values, to obtain the countless solution set of system.

In the future for simplicity, record/writing the equations of OZLP, we will consider them the linearly independent; in this case, the rank of system r will be equal to the number of equations m.

Thus, if the number of equations of OZLP r = m is less than the number of variables n, then the system of linear equations has countless solution set, i.e., value parts x_1, x_2, ..., x_n, are

nonnegative, then this means that OZLP does not have the permissible solution.

But if there are some solutions of systems (2.1), for which everything $x_1, x_2, \ldots, x_n$ are nonnegative (are shorter, "nonnegative solutions"), then each of them is admissible. Appears problem - to find among the permissible solutions the optimum, i.e., this solution

$$x_1, x_2, \ldots, x_n,$$

for which the linear function

$$L = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

is converted into the minimum.

Page 45.

In order to more distinct visualize the special feature/peculiarities of the solution of OZLP and different cases which can in this case be met, it is convenient to use geometric interpretation.

3. Geometric interpretation of the basic problem of linear programming.

Let us consider the case when the number of variables n to two

is more than the number of independent equations m, by which they must satisfy:

$$n-m=2.$$

Then as we already know that it is possible two of n of variables, let us say x_1 and x_2 to select as free, and the others m to make base and to express them through free. Let us assume that this is done. We will obtain m = n - 2 equations of the form:

$$\left.\begin{array}{l} x_3=\alpha_{31}x_1+\alpha_{32}x_2+\beta_3, \\ x_4=\alpha_{41}x_1+\alpha_{42}x_2+\beta_4, \\ \cdots\cdots\cdots\cdots \\ x_n=\alpha_{n1}x_1+\alpha_{n2}x_2+\beta_n. \end{array}\right\} \quad (3.1)$$

Let us give to the problem of linear programming geometric interpretation. Along the axes Ox_1 and Ox_2, it will plot/deposit the values of unrestricted variables x_1 x_2 (Fig. 2.2).

Since the variables x_1, x_2 must be nonnegative, the allowed values of unrestricted variables lie/rest only higher than the axis Ox_1 and more to the right axis Ox_2; let us note this by the shading, which designates the "permissible side" of each coordinate axis.

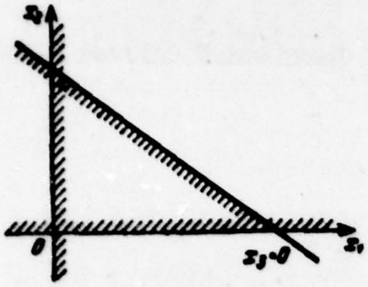


Fig. 2.2.

Page 46.

The remaining variables $x_3, x_4, \ldots, x_n$ also must be nonnegative, i.e., must be implemented the conditions:

$$\left.\begin{array}{l} x_3=\alpha_{31}x_1+\alpha_{32}x_2+\beta_3\geqslant 0, \\ x_4=\alpha_{41}x_1+\alpha_{42}x_2+\beta_4\geqslant 0, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ x_n=\alpha_{n1}x_1+\alpha_{n2}x_2+\beta_n\geqslant 0. \end{array}\right\} \quad (3.2)$$

Let us look how to depict these conditions geometrically. Let us take one of them, for example, the first:

$$x_3=\alpha_{31}x_1+\alpha_{32}x_2+\beta_3\geqslant 0.$$

Let us place value x_3 equal to our extreme value - zero. We will obtain the equation

$$\alpha_{31}x_1+\alpha_{32}x_2+\beta_3=0.$$

This - the equation of straight line. On this straight line $x_3 = 0$ (see Fig. 2.2); along one side from it $x_3 > 0$, on another $x_3 < 0$ (on which - this depends on the coefficients of equation). Let us note by shading TU the side of the straight line $x_3 = 0$, along which $x_3 > 0$.

Analogously let us construct everything remaining limiting lines: $x_4 = 0, \ldots, x_n = 0$ will note at each of them by shading the "permissible side", where the corresponding variable is more than zero (Fig. 2.3).

Thus, we will obtain n of straight lines: two axes of coordinates ($x_2 = 0$, $x_1 = 0$) and of n - 2 straight lines ($x_3 = 0, \ldots, x_n = 0$). Each of them determines the "permissible half-plane", which lies along its one side. The part of plane $x_1 0 x_2$, which belongs simultaneously to all half-planes, forms the domain of the permissible solutions (ODR). In Fig. 2.3 domain of the permissible solutions always represents by itself convex polygon.

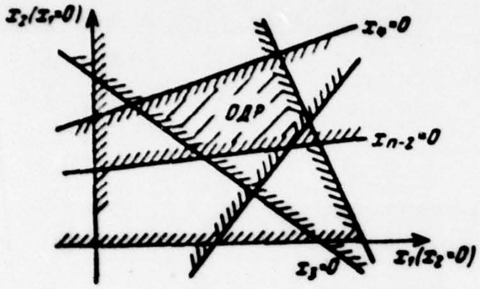


Fig. 2.3.

Page 47.

As is known, convex figure (Fig. 2.4) is called the figure, which possesses the following property: if two points A and B belong to this figure, then also entire segment AB also belongs to it.

Let us demonstrate that ODR is always convex figure. Let us assume the contrary: points A and B belong to ODR, but some point C between them does not belong (see Fig. 2.4). Then between point A, which belongs to ODR, and the point, belonging to it, without fail must pass some one of lines $x_k=0$: along one side of this direct/straight point satisfy condition $x_k \geqslant 0$, on another - they do not satisfy. Let this straight line intersect segment AB at some point D. Then points A and B, which lie along different sides from line, cannot simultaneously belong to ODR (for it all x_k are nonnegative), which contradicts condition.

Figures 2.3 shows such example when ODR exists, i.e., system of equations of OZLP has nonnegative solutions of system there does not exist. An example of this case is shown on fig. 2.5. It is real/actual, there is no domains, which lies along one and the same (shaded) side from all straight lines: $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, $x_4 = 0$, $x_5 = 0$; i.e. the conditions of the nonnegative character of variables contradict each other and the permissible solutions of OZLP do not exist.

Example 1. The problem of linear programming with family alternating/variable has m = 5 equation-limitations:

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7$$

we have m = 5 equations-limitations:

$$\left.\begin{aligned} x_1-x_2+x_3 &= 4; \\ 2x_1-x_2-x_3-x_4 &= -5; \\ x_1+x_2-x_5 &= -4; \\ x_2+x_6 &= 5; \\ 2x_1-2x_2-x_6+2x_7 &= 7. \end{aligned}\right| \qquad (3.3)$$

It is required to give its geometric interpretation and to construct ODR, if it exists.

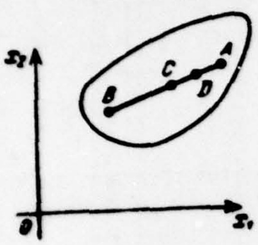


Fig. 2.4.

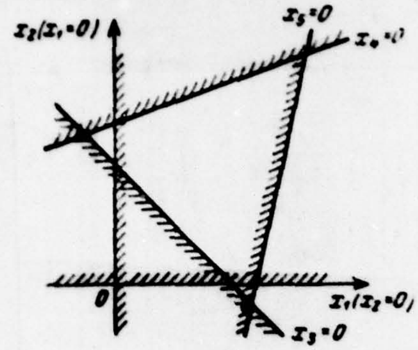


Fig. 2.5.

Page 48.

Solution. Let us select as unrestricted variables , for example, x_1 and x_2 is expressed by them the remaining (base) variables: x_3, x_4, x_5, x_6, x_7. From the first equation we have:

$$x_3 = -x_1 + x_2 + 4 \qquad (3.4)$$

From the third:

$$x_5 = x_1 + x_2 + 4.$$

From the fourth:

$$x_6 = -x_2 + 5. \qquad (3.5)$$

Substituting (3.4) in second equation (3.3) and (3.5) - in the latter and solving relative to x_4, x_7, we have:

$$x_4 = 3x_1 - 2x_2 + 1;$$
$$x_7 = -x_1 + \tfrac{1}{2}x_2 + 6.$$

The geometric interpretation of problem is represented in Fig. 2.6 (straight lines $x_1 = 0$, $x_2 = 0$ - coordinate axis; the remaining limiting straight lines $x_3 = 0$, $x_4 = 0$, $x_5 = 0$, $x_6 = 0$ and $x_7 = 0$; short shading marked the permissible half-planes).

As can be seen from the location of the direct/straight and noted half-planes, the permissible solutions for the examined problem exist; they fill

Thus, we considered a question concerning the existence of the domain of the permissible solutions of OZLP and (for case of $m = n-2$) gave to it geometric interpretation.

Now arises the question concerning determination from the number permissible of the optimum solution, i.e., such, which converts into the minimum the linear function

$$L = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n. \qquad (3.6)$$

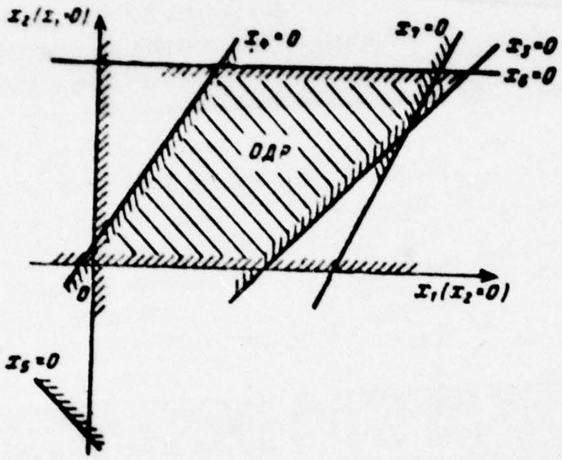


Fig. 2.6.

Page 49.

Let us give to this problem geometric interpretation, moreover again for the case when $m = n-2$ (i.e. number free variables is equal to 2, and number base m).

Let us assume that unrestricted variables they are again x_1, x_2, and base x_3, x_4, ..., x_n, the expressed through free by formulas (3.2). Let us substitute expressions (3.2) into formula (3.6), let us give similar terms and is expressed the linear function L all n of variables as linear function only of of two unrestricted variables: x_1 and x_2. We will obtain:

$$L = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2, \qquad (3.7)$$

where γ_0 - absolute term which in initial form of function L was not; now, during transition to the variables x_1, x_2, it could appear.

It is obvious, linear function (3.7) reaches the minimum at the same values x_1, x_2, that also the function

$$L'=\gamma_1 x_1+\gamma_2 x_2$$

without absolute term (linear form). Is real/actual, $L' = L - \gamma_0$, where γ_0 does not depend on x_1 and x_2, and, obviously, the minimums of that and other of functions, that differ on γ_0, are reached at one and the same values x_1, x_2.

Let us find these values, using geometric interpretation. Let us give L' certain constant value of C:

$$L'=\gamma_1 x_1+\gamma_2 x_2=C;$$

we will obtain equation of straight line on plane x_1Ox_2 (Fig. 2.7). The angular coefficient of this straight line is equal $-\gamma_1/\gamma_2$, and the segment, intercept/detached by it on axis Ox_2 (initial ordinate), is equal to C/γ_2. It is obvious, if we replace the constant C for certain other C_1, the angular coefficient of straight line will not

be changed; will be changed only initial ordinate, and straight line will move by in parallel to itself to the new position $L' = C_1$ (see Fig. 2.7).

Thus, to different values L' correspond different straight lines on plane but they all are parallel between themselves.

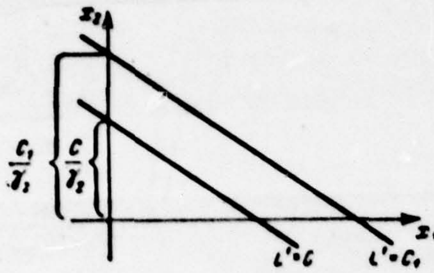


Fig. 2.7.

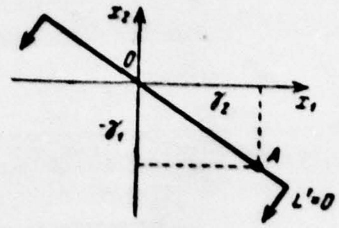


Fig. 2.8.

Page 50.

It is obvious, instead of all these straight lines sufficient to depict on plane one basic straight line, for example, $L' = 0$, and then it is possible to mentally move it in parallel to itself. During transferring of this straight line to one side L', it will grow, into another - to decrease.

Let us construct the basic straight line $L' = 0$ on plane x_1Ox_2 (Fig. 2.8). We know that its angular coefficient is equal - γ_1/γ_2: in order to construct straight line, passing through the origin of coordinates with angular coefficient γ_1/γ_2, let us plot along the axis of the abscissas of cuttings off γ_2, and along the axis of the ordinates of cuttings off $-\gamma_1$, and through point A with

such coordinates let us draw straight line. This there will be the basic straight line L' = 0.

Now there remains only to explain that to which side (in parallel itself) it is necessary to move this straight line so that value L' would decrease. In the case, shown on Fig. 2.8 (both coefficient γ_1 and γ_2 are positive) the direction of decrease L' - downward and to the left (this is shown by rifleman/pointers, directed from basic straight line to the side of decrease L'). With other signs of coefficients γ_1, γ_2, the direction of decrease varies. The cases of different directions of decrease are shown on Fig. 2.9, 2.10 and 2.11.

Thus, and direction the basic straight line L' = 0, and the direction of the decrease of the linear form L' are determined by values and the signs of the coefficients γ_1, γ_2 of unrestricted variables x_1, x_2 in expression L'.

Let us give now the geometric interpretation of the determination of the optimum solution of OZLP among permissible.

Let there be the domain of the permissible solutions ODR (Fig. 2.12) and the basic straight line L' = 0; known (is shown by rifleman/pointers) the direction of the decrease of the linear form L'.

During transferring basic straight line in the direction, indicated by rifleman/pointers, the linear form L' will decrease. it is obvious, the smallest value it will achieve, when straight line will pass through the extreme point ODR, outermost from the origin of coordinates in direction rifleman/gunner (in our case, point A). The coordinates of this point x_1^*, x_2^* determine the optimum solution of OZLP.

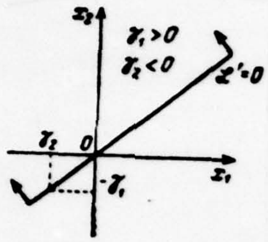


Fig. 2.9.

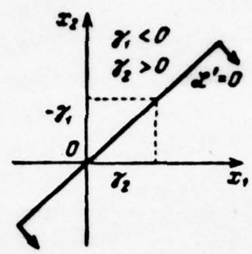


Fig. 2.10.

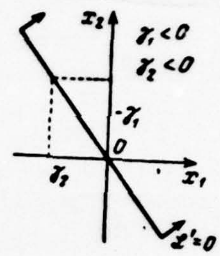


Fig. 2.11.

Page 51.

Knowing the optimum values of unrestricted variables x_1^*, x_2^*, it is possible tc find, substituting them in equations (3.2), and the optimum values of the base variables:

$$x_3^* = \alpha_{31} x_1^* + \alpha_{32} x_2^* + \beta_3,$$
$$x_4^* = \alpha_{41} x_1^* + \alpha_{42} x_2^* + \beta_4,$$
$$\cdots\cdots\cdots\cdots$$
$$x_n^* = \alpha_{n1} x_1^* + \alpha_{n2} x_2^* + \beta_n,$$

and also the optimum (minimum) value of the linear function L:

$$L_{min} = \gamma_0 + \gamma_1 x_1^* + \gamma_2 x_2^*. \qquad (3.8)$$

Thus if the number of independent equation-limitations, by which they must satisfy the variables x_1, x_2, ..., x_n, to two is less than the number of variables n (i.e. into OZLP figure two unrestricted variables and any number of base), the solution of OZLP can be

obtained by simple geometric construction.

Example 2. Under conditions of example 1 to find the optimum solution of OZLP, which rotates in the minimum the linear function of seven unknowns:

$$L = x_1 - x_2 + 2x_3 - x_4 - 3x_5 + x_6 - 2x_7. \tag{3.9}$$

Equation-limitation - the same as in example 1.

Solution. In example 1 of an equation-limitation (3.3) were solved relative to the base variables x_3, x_4, x_5, x_6, x_7 which were expressed through the free x_1 and x_2:

$$\left.\begin{aligned} x_3 &= -x_1 + x_2 + 4; \\ x_4 &= 3x_1 - 2x_2 + 1; \\ x_5 &= x_1 + x_2 + 4; \\ x_6 &= -x_2 + 5; \\ x_7 &= -x_1 + \tfrac{1}{2}x_2 + 6. \end{aligned}\right\} \tag{3.10}$$

Substituting these expressions in (3.9) and giving similar terms, we have:

$$L = -5x_1 - 2x_2 - 12. \tag{3.11}$$

Let us reproduce the domain of the permissible solutions, that previously constructed in Fig. 2.6 (see Fig. 2.13).

Reject/throwing absolute term in (3.11), we have

$$L' = -5x_1 - 2x_2.$$

We construct the basic straight line $L' = 0$.

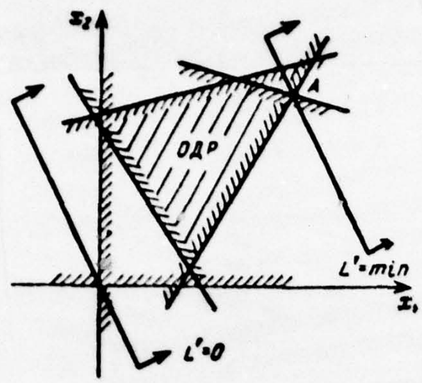


Fig. 2.12.

Page 52.

For this, we plot/deposit segments $\gamma_2 = -2$ along the axis of abscissas and $-\gamma_1 = 5$ along the axis of ordinates, we carry out through point B with coordinates (-2, 5) the straight line L' = 0 and note by rifleman/pointers the direction of decrease L'. Moving basic straight line in parallel to itself to the side of decrease L', the small value L' we will obtain at point A (outermost from beginning of coordinates in direction of arrow). The coordinates of this point x_1^*, x_2^* give the optimum solution of OZLP. In point A intersect two limiting straight lines: $x_6 = 0$ and $x_7 = 0$. Equalizing zero expressions for x_6 and x_7, we will obtain two equations:

$$\left.\begin{aligned} -x_2+5=0, \\ -x_1+\tfrac{1}{2}x_2+6=0. \end{aligned}\right\}$$

Solving them together, let us find $x_1^* = 8.5$; $x_2^* = 5$.

Substituting these values in (3.11), let us find the optimum values of the base variables:

$$x_3^* = 0,5; \quad x_4^* = 16,5; \quad x_5^* = 17,5.$$

As concerns x_6 and x_7, their optimum values are equal to zero: $x_6^* = 0$; $x_7^* = 0$.

Substituting the obtained optimum values x_1^* and x_2^* In linear function (3.11), let us find the minimum value (optimum) of the linear function L:

$$L = -5 \cdot 8,5 - 2 \cdot 5 - 12 = -64,5.$$

Thus, we learned to solve OZLP in the particular case of m = n - 2 with the help of geometric construction.

In spite of the fact that this construction is related to a special case, from it escape/ensue some overall considerations, which relate generally to the properties of the solution of OZLP.

Let us note noticed by us laws for case of n - m = 2.

1. Solution of OZLP, if it exists, cannot lie/rest at inside of domain of permissible solutions, but only on its boundary.

2. Solution of OZLP can be and not only (see Fig. 2.14). It is real/actual, if basic straight line is parallel to that side of the polygon of the permissible solutions where is reached minimum L', then it is achieved not at one point, but on an entire this side.

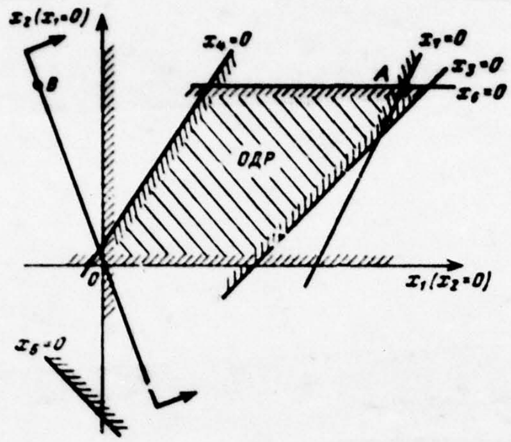


Fig. 2.13.

Page 53.

In this case OZLP has a countless multitude of optimum solutions.

3. OZLP can not have solution even in the case when there is ODR (Fig. 2.15). This occurs when in direction rifleman/gunner ODR it is not limited, i.e., in the domain of the permissible solutions, the linear function L is not limited from below. Moving basic straight line in direction rifleman/gunner, we will obtain increasingly smaller and smaller values L', and also, therefore, L.

4. Solution of OZLP, which minimizes function L (optimum

solution), always is achieved in one of apex/vertexes of polygon of permissible solutions (if it is reached on whole side, then it is achieved, also, in each of apex/vertexes through which passes this side). The solution, which lies at one of the apex/vertexes ODR, is called supporting/reference solution, and apex/vertex itself - by data points.

5. In order to find optimum solution, in principle is sufficient to sort out all the apex/vertexes ODR (data points) and to select of them their/that, where function L reaches minimum.

6. If number of unrestricted variables into OZLP is equal to 2, and number of base - m and solution of OZLP exists, then it always is reached in point where at least two of variables $x_1, x_2, \ldots, x_n$ are converted into zero. It is real/actual, at any data points intersect at least two of the limiting straight lines; can in it intersect more than two (see Fig. 2.16).

The case when in optimum solution they are converted into zero not two, but is more variables, it is called degenerated. Figures 2.16 shows the degenerate case when at point A, which corresponds to optimum solution, are converted into zero three variables: x_3, x_5 and x_6.

After considering in detail geometric interpretation for case of $m = n - 2$, let us turn to the case when the number of variables exceeds by 3 numbers of independent equation-limitations: $m = n-3$.

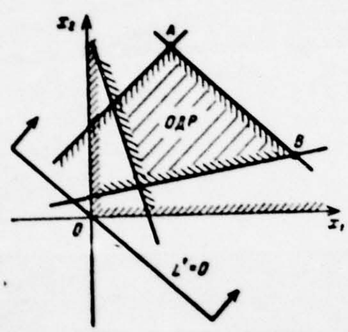


Fig. 2.14.

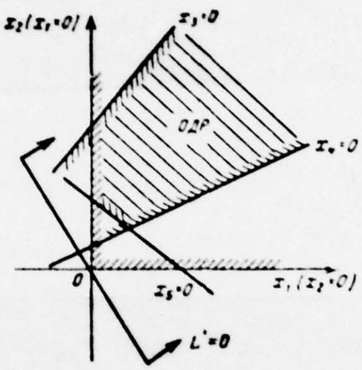


Fig. 2.15.

Page 54.

In this case unrestricted variables proves to be no longer two, but three (this will be x_1, x_2, x_3), and remaining m = n-3 of base variables they can be expressed through free:

$$\left.\begin{aligned} x_4 &= \alpha_{41}x_1 + \alpha_{42}x_2 + \alpha_{43}x_3 + \beta_4; \\ x_5 &= \alpha_{51}x_1 + \alpha_{52}x_2 + \alpha_{53}x_3 + \beta_5; \\ &\dots\dots\dots\dots\dots \\ x_n &= \alpha_{n1}x_1 + \alpha_{n2}x_2 + \alpha_{n3}x_3 + \beta_n. \end{aligned}\right\} \quad (3.12)$$

it is required to find such nonnegative values $x_1, x_2, \ldots, x_n$, which, satisfying equations (3.12), would simultaneously convert into the minimum the linear function of these variables:

$$L = c_1x_1 + c_2x_2 + \ldots + c_nx_n. \quad (3.13)$$

The geometric interpretation of this problem it is necessary to construct no longer on plane, but in space (Fig. 2.17). Each condition $x_k = 0$ for one of base variables x_k (k = 4, ..., n) will be geometrically depicted no longer straight line, but plane. Along one side from this plane $x_k > 0$, on another $x_k < 0$. The coordinate planes x_2Ox_3, x_1Ox_3 and x_1Ox_2 represent conditions $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ respectively. The domain of the permissible solutions (if it exists) it represents by itself the convex polyhedron, limited by these planes, i.e., the part of the space, for which are satisfied all conditions:

$$x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_3 \geqslant 0, \ldots, \quad x_n \geqslant 0.$$

Role "basic straight line" in this case will play "reference plane" whose equation L' = 0, where

$$L' = L - \gamma_0; \quad L = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3.$$

During transferring of this plane in parallel to itself to one side L' it will decrease, into another - to grow.

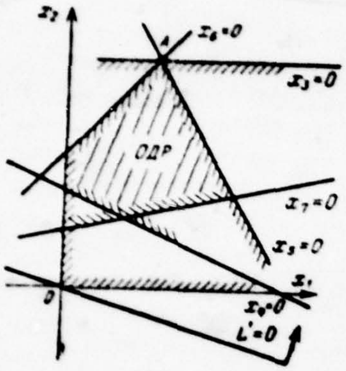


Fig. 2.16.

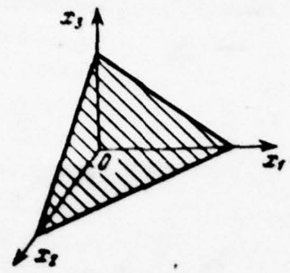


Fig. 2.17.

Page 55.

Point A, at which it is reached the optimum solution (if it exists), represents by itself the apex/vertex ODR which it is located further anything from the origin of coordinates, counting in the direction of decrease L'. Can render/show as with n - m = 2, that OZLP has countless solution set, either filling whole fin/edge or - whole face of polyhedron of the permissible solutions. The optimum solution x_1^*, x_2^*, x_3^* (if it exists) coincides from one of data points, i.e., the apex/vertexes of polyhedron, in which at least three variables $x_1, x_2, \ldots, x_n$ they are converted into zero.

To use geometric interpretation for the direct finding of solution even with n - m = 3 is difficult; with n - m = k > 3 this

will generally deduce us beyond the framework of three-dimensional space and geometric interpretation will lose clarity. However, the corresponding terminology can render/show convenient: it is possible to speak about the domain of the permissible solutions as to certain "super-polyhedron" in space k of measurements, limited m by "hyperplanes"; to optimum solution - as to one of the "apex/vertexes" of this polyhedron, to each "apex/vertex" - as to "data points", etc. By this geometric terminology it is possible, at will, to use or not to use. By us geometric interpretation was required for justifying the of following properties of the solution of OZLP at any values of the number of variables n and of the number of equations m < n:

1. Optimum solution, if it exists, lie/rests not inside, but on boundary of the domain of the permissible solutions, in one of data points, in each of which at least k of variables are converted into zero.

2. In order to find optimum solution, it is necessary, passing from one data points to another, to move in direction of decrease of linear function L, which it is required to minimize.

On these principles will be based the methods of the solution of OZLP which we is presented subsequently.

4. Problem of linear programming with limitation-inequalities. Transition to OZLP and conversely.

In practice of limitation in the problem of the linear programming frequently are given not by equations, but inequalities.

Let us show how it is possible to pass from problem with limitation-inequalities to the basic problem of linear programming.

Let there be the problem of linear programming with n by variables $x_1, x_2, \ldots, x_n$, in which the restrictions placed on variables, take the form of linear inequalities. In some of them inequality sign can be $\geqslant$, and others $\leqslant$ (second form is reduced to the first by a simple change in the sign of both parts).

Page 56.

Therefore let us assign all limitation-inequalities in the standard form:

Let us consider that all these inequalities are linearly independent (i.e. any of them it cannot be represented in the form of the linear combination of others).

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + b_1 \geqslant 0; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_2 \geqslant 0; \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + b_m \geqslant 0. \end{array}\right\} \qquad (4.1)$$

It is required to find such set of nonnegative values $x_1, x_2, \ldots, x_n$, which would satisfy inequalities (4.1), and, furthermore, would be converted into the minimum the linear function:

$$L=c_1x_1+c_2x_2+\ldots+c_nx_n. \tag{4.2}$$

From stated thus problem easily tc pass to the basic problem of linear programming. It is real/actual, let us introduce the designations:

$$\left.\begin{aligned} y_1 &= a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n+b_1, \\ y_2 &= a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n+b_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m &= a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n+b_m, \end{aligned}\right\} \tag{4.3}$$

where $y_1, y_2, \ldots, y_m$ - some new variables which we will call "additional". According to conditions (4.1), these additional variables just as $x_1, x_2, \ldots, x_n$, they must be nonnegative.

Thus, before us appears the problem of linear programming in the following setting: to find such nonnegative values of n + m of variables $x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m$, so that they would satisfy the system of equations (4.3) and simultaneously was converted into the minimum the linear function of these variables:

$$L=c_1x_1+c_2x_2+\ldots+c_nx_n.$$

As is evident, before us in pure form the basic problem of

linear programming (OZLP). Equations (4.3) are assigned in the form, already solved relative to base variables $y_1, y_2, \ldots, y_m$, which are expressed through unrestricted variables $x_1, x_2, \ldots, x_n$. The total quantity alternating/variable is equal to n + m of them n "of initial" and m "of additional". Function L is expressed only through the "initial" variables (coefficients of "additional" variables in it are equal to zero).

Thus, the problem of linear programming with limitations-inequalities have reduced we to the basic problem of linear programming, but with the large number of variables how it was initially in problem.

Page 57.

Example 1. There is a problem of linear programming with limitation-inequalities: to find the nonnegative values of the variables x_1, x_2, x_3, x_4, x_5, that satisfy the conditions

$$\left.\begin{aligned} 2x_1 - x_2 + 3x_3 &< 6, \\ x_3 - 3x_2 &< -1, \\ x_5 - 2x_4 + x_1 &> -1, \\ x_5 - x_1 &< 0 \end{aligned}\right\} \qquad (4.4)$$

and rotating in the minimum the linear function

$$L = x_1 - 2x_2 - 3x_3. \qquad (4.5)$$

It is required to lead this problem to the form of OZLP.

Solution. We lead inequalities (4.4) to the standard form:

$$\begin{aligned} -2x_1+x_2-3x_3+6 &> 0, \\ 3x_2-x_3-1 &> 0, \\ x_1-2x_4+x_5+1 &> 0, \\ x_1-x_5 &> 0. \end{aligned}$$

It is introduced the further variables:

$$\left.\begin{aligned} y_1 &= -2x_1+x_2-3x_3+6, \\ y_2 &= 3x_2-x_3-1, \\ y_3 &= x_1-2x_4+x_5+1, \\ y_4 &= x_1-x_5. \end{aligned}\right\} \qquad (4.6)$$

Problem is reduced to to find the nonnegative values of variables

$$x_1,\ x_2,\ x_3,\ x_4,\ x_5;\ y_1\ y_2,\ y_3,\ y_4,$$

satisfying equations (4.6) and rotating into the minimum linear function (4.5).

We showed, as from the problem of linear programming with limitation-inequalities it is possible to pass to problem with limitation-equalities (OZLP). Is always feasible reverse transition - from OZLP to problem with limitation-inequalities. If in the first case we increased the number of variables, then in the second case let us it reduce, removing base variables and leaving only free.

Example of 2. There is a problem of linear programming with limitation-equalities (OZLP):

$$\left.\begin{aligned} x_1+x_2 &= 1, \\ x_2-2x_3 &= -3, \\ x_3-x_4+x_5 &= 1 \end{aligned}\right\} \qquad (4.7)$$

and by the minimized function

$$L = -x_1 - x_2 + x_5. \tag{4.8}$$

It is required to register it as problem of linear programming with limitation-inequalities.

Solution. Since m = 3, n = 5, n - m = 2,, then let us select some two of the variables as free. Let us note that the variables x_1, x_2 as free cannot be chosen, since they are connected by the first of equations (4.7): value of one of them is completely determined by value another, and unrestricted variables must be independent variables.

Page 58.

On the same reason it is not possible as free to select the variables x_2, x_3 (from connects the second equation (4.7)). Let us select as the unrestricted variables x_1 and x_4 and it is expressed by them all others:

$$\left.\begin{aligned} x_2 &= -x_1 + 1, \\ x_3 &= -\tfrac{1}{2}x_1 + 2, \\ x_5 &= \tfrac{1}{2}x_1 + x_4 - 1. \end{aligned}\right\} \tag{4.9}$$

Since $x_2 > 0$, $x_3 > 0$, $x_5 > 0$, conditions (4.9) can be replaced by the inequalities:

$$\left.\begin{aligned} -x_1 \quad + 1 &> 0, \\ -\tfrac{1}{2}x_1 + 2 &> 0, \\ \tfrac{1}{2}x_1 + x_4 - 1 &> 0. \end{aligned}\right\} \qquad (4.10)$$

Let us pass in the expression of the linear function L to unrestricted variables x_1, x_4. Substituting in L for x_2 and x_5 of their expression (4.9), we will obtain:

$$\begin{gathered} L = -x_1 + x_1 - 1 + \tfrac{1}{2}x_1 + x_4 - 1 = \tfrac{1}{2}x_1 + x_4 - 2, \\ L' = \tfrac{1}{2}x_1 + x_4. \end{gathered} \qquad (4.11)$$

Thus, problem is reduced to the problem of linear programming with limitation-inequalities. Its geometric interpretation is shown on Fig. 2.18. The basic straight line L' = 0 is parallel to that side ODR where L' it reaches the minimum. Consequently, all points of section AB give optimum solution. Taking as the solution, for example, the coordinates of point A, we will obtain:

$$x_1^* = 0; \quad x_4^* = 1; \quad x_2^* = 1; \quad x_3^* = 2; \quad x_5^* = 0.$$

At such values of variables, the linear function L reaches the minimum, equal to

$$L_{min} = \tfrac{1}{2} \cdot 0 + 1 - 2 = -1.$$

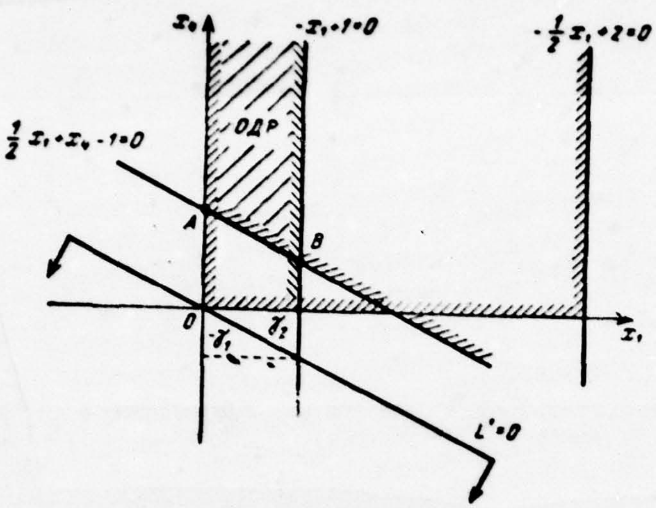


Fig. 2.18.

Page 59.

Thus, we can on arbitrariness pass from OZLP to the problem of linear programming with limitations by inequalities and vice versa. If we in the number of limitations of the problem eat both the equations and the inequalities, is recommended to produce unification and to pass in any uniform form, for example OZLP.

Example 3. Is examined the problem of linear programming with the variables x_1, x_2, x_3, x_4 and the limitations of the form

$$\left.\begin{aligned} x_1+x_2&=x_3+x_4,\\ x_1-x_2+x_3&<1,\\ x_2+x_3+x_4&>5. \end{aligned}\right\} \qquad (4.12)$$

Is minimized the function

$$L = x_1 - 2x_2 + x_3 - 3x_4. \tag{4.13}$$

It is required to give the problem to of OZLP.

$$\left.\begin{aligned} x_1 + x_2 - x_3 - x_4 &= 0, \\ x_1 - x_2 + x_3 + y_1 &= 1, \\ x_2 + x_3 + x_4 - y_2 &= 5. \end{aligned}\right\}$$

Solution. By the introduction of the additional variables y_1, y_2 let us lead conditions (4.12) to the form of OZLP:

Minimized function remains in the form (4.13).

5. Simplex method of the solution of the problem of linear programming.

The geometric interpretation, which we used during the solution of the problems of linear programming, ceases to be suitable for this purpose so on the number of unrestricted variables n-m > 3, and is difficult already with n-m = 3. For the determination of the solution of the problem of linear programming in the general case (with the arbitrary number of unrestricted variables) are applied not geometric, but computational methods. From them most universal is so-called simplex method.

The idea of the simplex method is relatively simple. Let there be in problem of linear programming n of variables and m of the independent linear limitations, assigned in the form of equations. We know that the optimum solution (if it exists) it is reached in one of data points (apex/vertexes ODR), where at least k = n-m of variables are equal to zero. Let us select some k of variables as free and it is expressed by them the others m of base variables. Let, for example, as free be selected first k = n-m alternating/variable $x_1, x_2, \ldots, x_k$, and the others m are expressed through them:

$$\left.\begin{aligned} x_{k+1} &= \alpha_{k+1,1} x_1 + \alpha_{k+1,2} x_2 + \ldots + \alpha_{k+1,k} x_k + \beta_{k+1}, \\ x_{k+2} &= \alpha_{k+2,1} x_1 + \alpha_{k+2,2} x_2 + \ldots + \alpha_{k+2,k} x_k + \beta_{k+2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= \alpha_{n,1} x_1 + \alpha_{n,2} x_2 + \ldots + \alpha_{n,k} x_k + \beta_n. \end{aligned}\right\} \quad (5.1)$$

Page 60.

Let us try, which will be, if we place all unrestricted variables $x_1, x_2, \ldots, x_k$ equal to zero:

$$x_1 = 0, \quad x_2 = 0, \ldots, \quad x_k = 0.$$

In this case, we will obtain:

$$x_{k+1} = \beta_{k+1}, \quad x_{k+2} = \beta_{k+2}, \quad \ldots, \quad x_n = \beta_n$$

This solution can be permissible or that not admitted. It is admissible, if all absolute terms $\beta_{k+1}, \beta_{k+2}, \ldots, \beta_n$ are nonnegative. Let us assume that this condition is satisfied. Then we obtained supporting/reference solution. But is it optimum? It can be yes, while it can be and no. In order to check this, it is expressed the minimized linear function L through unrestricted variables $x_1, x_2, \ldots, x_k$:

$$L = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_k x_k. \qquad (5.2)$$

It is obvious that when $x_1 = x_2 = \ldots = x_k = 0$ $L = \gamma_0$. Let us look, cannot we improve solution, i.e., decrease function L, increasing any of the variables $x_1, x_2, \ldots, x_k$ (to reduce them we not can, since they all are equal to zero, but the negative values of variables are not admitted). If all coefficients $\gamma_1, \gamma_2, \ldots, \gamma_k$ in formula (5.2) are positive, then, increasing some one of the variables $x_1, x_2, \ldots, x_k$ over zero, we cannot decrease L; consequently, the found by us supporting/reference solution is optimum. But if among coefficients $\gamma_1, \gamma_2, \ldots, \gamma_k$ in formula (5.2) there is negative, then, increasing some of the variables $x_1, x_2, \ldots, x_k$, namely - those, the coefficients of which are negative, we can improve solution, i.e., decrease L.

Let, for example, the coefficient γ_1 in formula (5.2) be negative. That means that there is sense to increase x_1, i.e., to pass from this supporting/reference solution to other, where the

variable x_1 is not equal to zero, but instead of it is equal to zero some another. Increase x_1 "is useful" for the linear function L, it makes it it is smaller. However, to increase x_1 is necessary carefully, so as to would not become negative other variables $x_{h+1}, x_{h+2}, \ldots, x_n$, expressed through the unrestricted variables, in particular, through x_1 by formulas (5.1).

Let us look, it is dangerous for variables $x_{h+1}, x_{h+2}, \ldots, x_n$ increase x_1, i.e., can it do them negative? Yes, it is dangerous, if the coefficient of x_1 in the appropriate equation is negative. If among equations (5.1) there is no equation with the negative coefficient of x_1, then value x_1 can be increased boundless, and, which means, that the linear function L is not limited from below and optimum solution of OZLP does not exist.

Let us assume that this not so and that among equations (5.1) there is such, in which the coefficient of x_1 is negative. For variables, that stand in the left sides of these equations, increase x_1 dangerously - it can do them negative.

Page 61.

Let us take one of such variables x_l and let us look, to what extent it is possible all the same to increase x_1 until variable x_l

does become negative? Let us write out the l equation from system (5.1):

$$x_l = \alpha_{l1} x_1 + \alpha_{l2} x_2 + \ldots + \alpha_{lk} x_k + \beta_l.$$

Here absolute term $\beta_l \geqslant 0$, and coefficient α_{l1} is negative. It is easy to comprehend that if we leave $x_2 = \ldots = x_k = 0$, then x_1 we can increase only to the value equal to $-\beta_l/\alpha_{l1}$, and with further increase x_1, variable x_l will become negative.

Let us select ~~by~~ them from variables $x_{k+1}, \ldots, x_n$, which earlier than all will become zero with an increase x_1, i.e., for which the value $-\beta_l/\alpha_{l1}$ is smaller anything. Let such "most threatened" the variable will be x_r. Then has sense to re-solve a system of equations (5.1) relative to other base variables, removing from the number of unrestricted variables x_1 and after transferring instead of it into the group of unrestricted variables x_r. It is real/actual, we wish to pass from the supporting/reference solution, given by equalities $x_1 = x_2 = \ldots = x_k = 0$, to the supporting/reference solution in which already $x_1 \neq 0$, and $x_2 = \ldots = x_k = x_r = 0$. The first supporting/reference solution we obtained, after placing equal to zero all previous unrestricted variables $x_1, x_2, \ldots, x_k$; the second we we will obtain, if we will turn into zero all new unrestricted variables $x_2, \ldots, x_k, x_r$. Base variables in this case they will be $x_1, x_{k+1}, \ldots, x_{r-1}, x_{r+1}, \ldots, x_n$.

Let us assume that equations of type (5.1) for the new set of base and unrestricted variables are comprised. Then it is possible to express by new free variable and the linear function L. If all coefficients of variables in this formula are positive, then we found the optimum solution: it will be obtained, if all unrestricted variables are assumed equal to zero. If among the coefficients of variables there is negative, then the procedure of an improvement in the solution is continued: system again is re-solved relative to other base variables, and so on until is found the optimum solution, which rotates function L in the minimum.

Let us observe the described procedure of a gradual improvement in the solution of OZLP based on specific example.

Example. There is a problem of linear programming with limitation-inequalities:

$$\left.\begin{aligned} -5x_1-x_2+2x_3 &< 2, \\ -x_1+x_3+x_4 &< 5, \\ -3x_1+5x_4 &< 7. \end{aligned}\right\} \qquad (5.3)$$

It is required to minimize the linear function

$$L=5x_1-2x_3.$$

Solution. Reducing inequalities to standard form (≥ 0) and introducing the additional variables y_1, y_2, y_3, we pass to

condition-equalities:

$$\left.\begin{aligned} y_1 &= 5x_1 + x_2 - 2x_3 + 2, \\ y_2 &= x_1 - x_3 - x_4 + 5, \\ y_3 &= 3x_1 - 5x_4 + 7. \end{aligned}\right\} \qquad (5.4)$$

Page 62.

Number alternating/variable $n = 7$ by 4 exceeds the number of equations $m = 3$. That means that four variables they can be selected as free.

Let us try to select as unrestricted variables x_1, x_2, x_3, x_4 and to place them equal to zero. In this case, we will immediately obtain the supporting/reference solution: $x_1 = x_2 = x_3 = x_4 = 0$; $y_1 = 2$; $y_2 = 5$; $y_3 = 7$.

At these values alternating/variable $L = 0$.

Let us look, is this solution optimum? No ! Because in the expression of the linear function L the coefficient of x_3 is negative. That means increasing x_3, it is possible to decrease L.

Let us try to increase x_3. Let us observe according to equations (5.4), is dangerous this for other variables? Yes, it is dangerous for y_1 and y_2 - in both these equations the variable x_3 enters with

negative coefficient, which means, that with increase x_3, the corresponding variables y_1 and y_2 can become negative.

Let us look, which of these variables y_1 or y_2 is that most" threatened", which more earlily will become zero with increase x_3. It is obvious, y_1: it will become equal to zero with $x_3 = 1$, and value y_2 - only with $x_3 = 5$.

Therefore it is selected alternating/variable y_1 and it is introduced it into the number of free instead of x_3. In order "to re-solve" system (5.4) relative to x_3, y_2, y_3, let us act by the following manner. It is solved first equation (5.4) relative to the new base variable x_3:

$$x_3 = {}^2/_3 x_1 + {}^1/_3 x_2 - {}^1/_3 y_1 + 1.$$

This expression let us substitute for x_3 in the second equation; we will obtain

$$y_2 = -{}^2/_3 x_1 - {}^1/_3 x_2 + {}^1/_3 y_1 - x_4 + 4.$$

As concerns third equation, it, as not not containing x_3, will not be changed. Thus, we led system (5.4) to the form:

$$\left.\begin{aligned} x_3 &= {}^2/_3 x_1 + {}^1/_3 x_2 - {}^1/_3 y_1 + 1, \\ y_2 &= -{}^2/_3 x_1 - {}^1/_3 x_2 + {}^1/_3 y_1 - x_4 + 4, \\ y_3 &= 3x_1 - 5x_4 + 7 \end{aligned}\right\} \qquad (5.5)$$

with unrestricted variables x_1, x_2, y_1, x_4 and base x_3, y_2, y_3.

It is expressed the linear function L through the new unrestricted variables:

$$L = 5x_1 - 5x_1 - x_2 + y_1 - 2,$$

or

$$L = -x_2 + y_1 - 2. \tag{5.6}$$

Let us place now unrestricted variables equal to zero. The linear function L will become equal to -2. This it is already better than previous value L = 0. But is this solution optimum? Still no, since the coefficient of x_2 in expression (5.6) is negative. Thus, let us increase x_2. Let us look, for which of variables, that stand in the left sides of system (5.5), this can be "dangerously". Only for y_2 (in the first equation x_2 it enters with positive coefficient, but in the third in no way it enters).

Thus, it is exchanged by places alternating/variable x_2 and y_2 - the first let us deduce from the number of free, and the second - let us introduce. For this, is solved second equation (5.5) relative to x_2 and let us substitute this x_2 into the first equation. We will obtain one additional form of system (5.4):

$$\left.\begin{aligned} &x_2 = x_1 - y_2 - x_4 + 5, \\ &x_3 = -3x_1 - 2y_2 + y_1 - 2x_4 + 8; \\ &y_3 = 3x_1 - 5x_4 + 7. \end{aligned}\right\} \tag{5.7}$$

Is expressed L through the new unrestricted variables:

$$L = 3x_1 + 2y_2 - y_1 + 2x_4 - 8 + y_1 - 2,$$

or

$$L = 3x_1 + 2y_2 + 2x_4 - 10. \quad (5.8)$$

Page 63.

Set/assuming $x_1 = y_1 = y_2 = x_4 = 0$, we will obtain

$$L = -10$$

Is this solution optimum? This time - yes, since the coefficients of all unrestricted variables in expression (5.8) are nonnegative.

Thus, the optimum solution of OZLP is found:

$$x_1^* = 0, \quad x_2^* = 8; \quad x_3^* = 5; \quad x_4^* = 0; \quad y_1^* = 0; \quad y_2^* = 0; \quad y_3^* = 7.$$

At such values of variables, the linear function L takes the minimum value:

$$L_{min} = -10.$$

Let us note that in the examined example of us it was not necessary to seek the supporting/reference solution: it immediately was obtained, when we placed unrestricted variables equal to zero. This is explained by the fact that in equations (5.4) all the absolute terms were nonnegative and, which means, that the first hitting solution render/showed supporting/reference. If this render/shows not then, it will be possible to arrive at supporting/reference solution with the help of the same procedure of the interchange some base and unrestricted variables, re-solving of equation until absolute terms become nonnegative. As this is made, we will see subsequently (see §7).

6. Tabular algorithm of the replacement of base variables.

The procedure of "re-solving" of the system of equation-limitations OZLP of relatively new base variables can be substantially simplified, if it are formalized and are reduced to filling of standard tables along the specific system of the rules (it is shorter, to algorithm). This algorithm we will demonstrate based on the specific example (in its validity for any general case reader may be convinced independently).

Let us consider the system of five equation-limitations:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + b_1, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + b_2, \\ y_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + b_3, \\ y_4 &= a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + b_4, \\ y_5 &= a_{51}x_1 + a_{52}x_2 + a_{53}x_3 + a_{54}x_4 + b_5 \end{aligned}$$

with four unrestricted variables: x_1, x_2, x_3, x_4. Let we need to deduce from the number of free any variable, for example x_2, and to transfer it into base, but instead of it to introduce into the number of free some base variable, let us say y_3; it is shorter, we wish to interchange by the places of the variables x_2 and y_3. This replacement we will symbolically designate

$$x_2 \leftrightarrow y_3.$$

Page 64.

Let us look, which actions must be for this carried out.

Generally, it would be possible for each new system of equations to carry out re-solving anew, i.e., for replacement $x_2 \leftrightarrow y_3$, we would take in third equation (6.1) term $a_{32}x_2$, containing x_2, (let us name it the "solving term"; it goes without saying that we assume $a_{32} \neq 0$), they would transfer it into left side, and y_3 - into right; they would solve equation relative to x_2 and would substitute expression

for x_2 in all the remaining equations. Procedure sufficiently bulky, requiring the stressed attention; during its fulfillment is easy to be mistaken (especially with the large number of equations). However, since here each time it is necessary to make the same operations, then they is sufficient to fulfill one time in general form and to deduce the rules of the conversions, which then can be applied automatically. These rules, which realize "re-solving" of system, are conveniently realized in the form of tabular algorithm.

So that this algorithm would be simpler and more easily it was memorized, expedient to preliminarily somewhat convert system of equations (6.1), representing their right sides as differences between the absolute terms and the sum of the others:

$$\left.\begin{aligned}
y_1 &= b_1 - (-a_{11}x_1 - a_{12}x_2 - a_{13}x_3 - a_{14}x_4),\\
y_2 &= b_2 - (-a_{21}x_1 - a_{22}x_2 - a_{23}x_3 - a_{24}x_4),\\
y_3 &= b_3 - (-a_{31}x_1 - a_{32}x_2 - a_{33}x_3 - a_{34}x_4),\\
y_4 &= b_4 - (-a_{41}x_1 - a_{42}x_2 - a_{43}x_3 - a_{44}x_4),\\
y_5 &= b_5 - (-a_{51}x_1 - a_{52}x_2 - a_{53}x_3 - a_{54}x_4).
\end{aligned}\right\} \qquad (6.1)$$

Designating

$$-a_{11} = \alpha_{11}; \quad -a_{12} = \alpha_{12}; \; \dots ; \quad -a_{54} = \alpha_{54},$$

we will obtain:

$$\left.\begin{aligned}
y_1 &= b_1 - (\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{14}x_4),\\
y_2 &= b_2 - (\alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23}x_3 + \alpha_{24}x_4),\\
y_3 &= b_3 - (\alpha_{31}x_1 + \alpha_{32}x_2 + \alpha_{33}x_3 + \alpha_{34}x_4),\\
y_4 &= b_4 - (\alpha_{41}x_1 + \alpha_{42}x_2 + \alpha_{43}x_3 + \alpha_{44}x_4),\\
y_5 &= b_5 - (\alpha_{51}x_1 + \alpha_{52}x_2 + \alpha_{53}x_3 + \alpha_{54}x_4).
\end{aligned}\right\} \qquad (6.2)$$

The form of writing of equations (6.2) we will call standard.

It is obvious, instead of completely record/writing of equations (6.2), it is possible to be bounded to filling of the standard table where are shown only absolute terms and the coefficients of variables. The first column of table we will weigh out under absolute terms, the second, third, fourth and the fifth - under the coefficients of the variables x_1, x_2, x_3, x_4 in standard form (6.2). Standard table for system (6.2) is given in Table 6.1.

Let us now visualize that we wish to replace $x_2 \leftrightarrow y_3$, i.e., to transfer the variable x_2 into the number of base, and alternating/variable y_3 - into the number of free.

Page 65.

Let us isolate in the standard table the solving cell/element α_{32} (let us encircle by its small circle); it is isolated also by heavy lines row and column, in which stands the solving cell/element. This

row and this column we will call the solving row and solving column (see Table 6.2).

Implementing operation $x_2 \leftrightarrow y_3$, we wish in the solving row to place the variable y_3, and in the solving column - alternating/variable x_2 (this is noted in table next to row and column).

Let us find the coefficients which will have to place in table after exchange $x_2 \leftrightarrow y_3$. Let us begin from the transformation of the solving row. Solving third equation (6.2) relative to x_2, we will obtain:

$$x_2 = \frac{b_3}{\alpha_{32}} - \left(\frac{\alpha_{31}}{\alpha_{32}} x_1 + \frac{1}{\alpha_{32}} y_3 + \frac{\alpha_{33}}{\alpha_{32}} x_3 + \frac{\alpha_{34}}{\alpha_{32}} x_4\right). \qquad (6.3)$$

Thus, the converted cell/elements of the solving row are found. Let us comprise the rule of the transformation of remaining rows.

Table 6.1.

	Свободный член [1]	x_1	x_2	x_3	x_4
y_1	b_1	α_{11}	α_{12}	α_{13}	α_{14}
y_2	b_2	α_{21}	α_{22}	α_{23}	α_{24}
y_3	b_3	α_{31}	α_{32}	α_{33}	α_{34}
y_4	b_4	α_{41}	α_{42}	α_{43}	α_{44}
y_5	b_5	α_{51}	α_{52}	α_{53}	α_{54}

Key: (1). Absolute term.

Table 6.2.

				$y_3 \leftrightarrow$		
		Свободный член [1]	x_1	x_2	x_3	x_4
	y_1	b_1	α_{11}	α_{12}	α_{13}	α_{14}
	y_2	b_2	α_{21}	α_{22}	α_{23}	α_{24}
$x_2 \leftrightarrow$	y_3	b_3	α_{31}	α_{32}	α_{33}	α_{34}
	y_4	b_4	α_{41}	α_{42}	α_{43}	α_{44}
	y_5	b_5	α_{51}	α_{52}	α_{53}	α_{54}

Key: (1). Absolute term.

Page 56.

For this, let us substitute into first equation (6.2) instead of x_2

its expression (6.3). After bringing of similar terms, we will obtain

$$y_1=\left(b_1-\frac{a_{12}b_3}{a_{32}}\right)-\left[\left(a_{11}-\frac{a_{12}a_{31}}{a_{32}}\right)x_1-\left(\frac{a_{12}}{a_{32}}\right)y_3+\right.$$
$$\left.+\left(a_{13}-\frac{a_{12}a_{33}}{a_{32}}\right)x_3+\left(a_{14}-\frac{a_{12}a_{34}}{a_{32}}\right)x_4\right].$$

It is not difficult to ascertain that by completely analogous form are transformed all the remaining rows. As a result we will obtain the converted table (see Table 6.3), in which operation $x_2 \leftrightarrow y_3$ is already completed.

After considering Table 6.3, we can so formulate the translation algorithm of the coefficients of standard table.

1. Solving cell/element is substituted by reverse to it value.

2. All remaining cell/elements of solving row are divided into solving cell/element.

3. All cell/elements of solving column (except most solving cell/element) reverse sign and this is done by solving cell/element.

4. Each of remaining cell/elements undergoes following transformation: to it is adjoined product of cell/element, which stood in previous solving row on the same place in order (i.e. in the same column), to cell/element, which stands in new solving column on appropriate place (i.e. in the same row, as our cell/element).

Table 6.3.

	(1) Свободный член	x_1	y_3	x_3	x_4
y_1	$b_1-\frac{a_{12}\ b_3}{a_{32}}$	$a_{11}-\frac{a_{12}\ a_{31}}{a_{32}}$	$-\frac{a_{12}}{a_{32}}$	$a_{13}-\frac{a_{12}\ a_{33}}{a_{32}}$	$a_{14}-\frac{a_{12}\ a_{34}}{a_{32}}$
y_2	$b_2-\frac{a_{22}\ b_3}{a_{32}}$	$a_{21}-\frac{a_{22}\ a_{31}}{a_{32}}$	$-\frac{a_{22}}{a_{32}}$	$a_{23}-\frac{a_{22}\ a_{33}}{a_{32}}$	$a_{24}-\frac{a_{22}\ a_{34}}{a_{32}}$
x_2	$\frac{b_3}{a_{32}}$	$\frac{a_{31}}{a_{32}}$	$\frac{1}{a_{32}}$	$\frac{a_{33}}{a_{32}}$	$\frac{a_{34}}{a_{32}}$
y_4	$b_4-\frac{a_{42}\ b_3}{a_{32}}$	$a_{41}-\frac{a_{42}\ a_{31}}{a_{32}}$	$-\frac{a_{42}}{a_{32}}$	$a_{43}-\frac{a_{42}\ a_{33}}{a_{32}}$	$a_{44}-\frac{a_{42}\ a_{34}}{a_{32}}$
y_5	$b_5-\frac{a_{52}\ b_3}{a_{32}}$	$a_{51}-\frac{a_{52}\ a_{31}}{a_{32}}$	$-\frac{a_{52}}{a_{32}}$	$a_{53}-\frac{a_{52}\ a_{33}}{a_{32}}$	$a_{54}-\frac{a_{52}\ a_{34}}{a_{32}}$

Key: (1). Absolute term.

Page 67.

Last/latter rule can in the first reading seem not by entirely clear; let us show how it is used at least based on the example of the cell/element, which stands in the first row and the second column Table 6.3. New cell/element that stand in the first row and the second column Table 6.3. New cell/element is equal to previous (α_{11}) plus the product of the previous cell/element of the solving row α_{31}, which stands in the same column, that α_{11}, and the new cell/element of solving column ($-\alpha_{12}/\alpha_{32}$), which stands in the same row, as the converted cell/element.

It is not difficult to ascertain that the formulated rules of the transformation of standard table are valid for any number of equations and unrestricted variables and for any replacement $x_j \leftrightarrow y_i$.

The transformation of standard table during replacement $x_j \leftrightarrow y_i$ is convenient to produce, implementing all the auxiliary calculations here, in table, for which is separated the lower part of each nucleus.

The translation algorithm $x_j \leftrightarrow y_i$ of standard table is reduced in this case to following operations.

1. To isolate in table solving cell/element α_{ij}. To compute its reciprocal value $\lambda = 1/\alpha_{ij}$ and to register in the lower part of the same nucleus (in right lower to angle).

2. All cell/elements of solving row (except very α_{ij}) to multiply on λ; result to register in lower part of the same nucleus.

3. All cell/elements of solving column (except very α_{ij}) to multiply on $-\lambda$; result to register in lower part of the same nucleus.

4. To emphasize (or to isolate in another manner) in solving row all upper numbers (previous cell/elements), with exception/elimination of most solving cell/element of nucleus, and in solving column - all lower numbers (new cell/elements), with exception/elimination of most solving cell/element.

5. For each of cell/elements, which do not belong either to solving row or to that solving column, to register into lower part of nucleus product of isolated numbers, which stand in the same column and in the same row, as this cell/element.

6. To rewrite table, after replacing:

- x_j on y_i conversely,

- cell/elements of solving row and column - by numbers, which stand in lower parts of the same nuclei,

- each of remaining cell/elements - by sum of numbers, which stand in upper and lower part of the same nucleus.

Example 1. In to system of equations

$$\left.\begin{aligned} y_1 &= x_1 - x_2 + 2x_3 - 5, \\ y_2 &= 2x_1 - x_2 + 1, \\ y_3 &= 2x_2 - x_3 - 1, \\ y_4 &= -x_1 - x_3 + 2 \end{aligned}\right\} \qquad (6.4)$$

to replace $x_1 \leftrightarrow y_2$, i.e., to deduce from the number of unrestricted variables x_1 and instead of it to introduce y_2.

Page 68.

Solution. We record/write equations (6.3) in the form of the standard table (see tables 6.4), leaving in the lower part of each nucleus of sufficiently vacant place.

Is isolated by small circle the solving cell/element - 2 and heavy lines - solving row and column. We compute $\lambda = -1/2$. Auxiliary records let us keep in right lower to the angle of the nucleus (see Table 6.5).

Let us fill, according to point/items 1, 2 and 3 algorithms, the lower parts of the nuclei of those solving the rows also of column.

Let us isolate, after surrounding them by the framework, the upper numbers of solving row and the lower numbers of solving chair (except the most solving nucleus).

Further we already can fill all the remaining lower parts of the nuclei, multiplying the corresponding to them isolated numbers, which stand in the solving row and the solving column on the same places

that this nucleus (see Table 6.6).

We finish conversicn, for which rewrite table 6.6, substituting x_1 by y_2, the cell/elements of the solving row and column - by lower numbers of the same nuclei, and remaining cell/elements - by sum of the upper and lower numbers (see Table 6.7).

Thus, mu learned with the help of tabular algorithm to accomplish in equation-limitations any replacement $x_j \leftrightarrow y_i$.

Let us recall that in the problem of linear programming, besides equation-limitations, there exists even the linear function

$$L = c_0 + c_1 x_1 + c_2 x_2 + \ldots + c_j x_j + \ldots + c_n x_n,$$

which must be minimized. If this function is expressed through previous unrestricted variables $x_1, x_2, \ldots, x_n$, then, obviously, after replacement $x_j \leftrightarrow y_i$ it must be expressed by new unrestricted variables $x_1, x_2, \ldots, x_{j+1}, y_i, x_{j+1}, \ldots, x_n$. It is not difficult to ascertain that for this can be used the same algorithm, as for the transformation of any row of standard table. It is real/actual, leading L to the standard form

$$L = c_0 - (\gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_n x_n),$$

where $\gamma_1 = -c_1; \gamma_2 = -c_2; \ldots; \gamma_n = -c_n$, we we will obtain one

additional row (additioral) of the standard table which differs from the others only in terms of the fact that in it never is chosen the solving cell/element.

Tables 6.4.

$y_2 \leftrightarrow$ (above column x_1)

	Свободный член (1)	x_1	x_2	x_3
y_1	-5	-1	1	-2
$x_1 \leftrightarrow$ y_2	1	-2	1	0
y_3	-1	0	-2	1
y_4	2	1	0	1

Key: (1). Absolute term.

Page 69.

Tables 6.5.

$y_2 \updownarrow$ (above column x_1)

	Свободный член (1)	x_1	x_2	x_3
y_1	-5	-1; $-\frac{1}{2}$	1	-2
$x_1 \leftrightarrow$ y_2	1; $-\frac{1}{2}$	-2; $-\frac{1}{2}$	1; $-\frac{1}{2}$	0; 0
y_3	-1	0; 0	-2	1
y_4	2	1; $\frac{1}{2}$	0	1

Key: (1). Absolute term.

Tables 6.6.

		y_2 ↔		
	Свободный член (1)	x_1	x_2	x_3
y_1	-5 $-\frac{1}{2}$	-1 $-\frac{1}{2}$	1 $-\frac{1}{2}$	-2 0
x_1 ↔ y_2	1 $-\frac{1}{2}$	-2 $-\frac{1}{2}$	1 $-\frac{1}{2}$	0 0
y_3	-1 0	0 0	-2 0	1 0
y_4	2 $\frac{1}{2}$	1 $\frac{1}{2}$	0 $\frac{1}{2}$	1 0

Key: (1). Absolute term.

Tables 6.7.

	Свободный член (1)	y_2	x_2	x_3
y_1	$-\frac{11}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-2
x_1	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	0
y_3	-1	0	-2	1
y_4	$\frac{5}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1

Key: (1). Absolute term.

Page 70.

Example of 2. To do a replacement $x_1 \leftrightarrow y_2$ in system of equations

$$\left.\begin{aligned} y_1 &= x_1 - x_2 + x_3 - 1, \\ y_2 &= \tfrac{1}{2} x_1 - x_3 - 3, \\ y_3 &= 3x_2 - 2x_3 \end{aligned}\right\} \qquad (6.5)$$

and in the linear function

$$L = -x_1 + 2x_2 - x_3 + 1.$$

Solution. Let us fill the standard table, in upper row of which we place the linear function L (see Table 6.8).

~~Table 6.~~8.

TABLE 6.8.

	Свободный член (1)	$y_2 \updownarrow x_1$	x_2	x_3
L	1	1	-2	1
y_1	-1	-1	1	-1
$x_1 \leftrightarrow y_2$	-3	$-\frac{1}{2}$	0	1
y_3	0	0	-3	2

Key: (1). Absolute term.

Tables 6.9.

	Свободный член (1)	$y_2 \updownarrow x_1$	x_2	x_3
L	1 / -6	1 / $\boxed{2}$	-2 / 0	1 / 2
y_1	-1 / 6	-1 / $\boxed{-2}$	1 / 0	-1 / -2
$x_1 \leftrightarrow y_2$	$\boxed{-3}$ / 6	$-\frac{1}{2}$ / -2	$\boxed{0}$ / 0	$\boxed{1}$ / -2
y_3	0 / 0	0 / $\boxed{0}$	-3 / 0	2 / 0

Key: (1). Absolute term.

TABLE 6.10.

	Свободный член (1)	y_2	x_2	x_3
L	-5	2	-2	3
y_1	5	-2	1	-3
x_1	6	-2	0	-2
y_3	0	0	-3	2

Key: (1) absolute term.

Page 71.

For the execution of replacement $x_1 \leftrightarrow y_1$ in the same table let us make the further calculations (see Table 6.9).

By replacement $x_1 \leftrightarrow y_1$ table is reduced to the form (table 6.10).

With the help of the tabular algorithm of the exchange of alternating/variable of equations of OZLP, it is possible to solve any problem of linear programming or to ascertain that it does not have solution.

The determination of the solution of each problem of linear programming falls into two stage:

1) seeking the supporting/reference solution;

2) finding the optimum solution, which minimizes the linear function L.

In the process of the first stage incidentally is clarified, does have generally this problem the permissible (nonnegative) solutions; if yes, then is located the supporting/reference solution for which all unrestricted variables are equal to zero, and everything base are nonnegative.

In the process of the second stage incidentally is clarified, is limited from below the minimized function L; if no, then of the optimum solution there does not exist. If yes, then it is located after one or the other number of replacements $x_j \leftrightarrow y_i$.

Both stage solutions of OZLP are conveniently implemented with the help of the described translation algorithm of standard tables.

Page 72.

7. finding the supporting/reference solution of the basic problem of linear programming.

Let there be OZLP with limitation-equalities, registered in the standard form:

$$\left.\begin{array}{l} y_1 = b_1 - (\alpha_{11} x_1 + \alpha_{12} x_2 + \dots + \alpha_{1n} x_n), \\ y_2 = b_2 - (\alpha_{21} x_1 + \alpha_{22} x_2 + \dots + \alpha_{2n} x_n), \\ \dots\dots\dots\dots\dots\dots \\ y_m = b_m - (\alpha_{m1} x_1 + \alpha_{m2} x_2 + \dots + \alpha_{mn} x_n), \end{array}\right\} \quad (7.1)$$

those solved relative to base variables $y_1, y_2, \dots, y_m$, which are expressed through unrestricted variables $x_1, x_2, \dots, x_n$. In each

apex/vertex of ODR (supporting/reference solution) at least n of variables must be converted into zero. Let us try to obtain the supporting/reference solution, set/assuming in formulas (7.1) all unrestricted variables equal to zero.

We have:

$$x_1 = x_2 = \ldots = x_n = 0; \quad y_1 = b_1;\; y_2 = b_2; \ldots;\; y_m = b_m. \tag{7.2}$$

If all absolute terms $b_1, b_2, \ldots, b_m$ in equations (7.1) are nonnegative, this means that the supporting/reference solution is already obtained; this case us does not interest. Let us consider the case when among absolute terms $b_1, b_2, \ldots, b_m$ there is negative. This means that solution (7.2) is not reference - it not at all admissibly, and the supporting/reference solution still is in prospect to find. For this, we will step by step transpose base and unrestricted variables in equations (7.1) until we arrive at the supporting/reference solution or will not ascertain that it do not exist. The latter occurs in the case when system of equations (7.1) is incompatible with the inequalities

$$x_1 \geqslant 0,\; x_2 \geqslant 0, \ldots,\; x_n > 0,\; y_1 > 0, \ldots,\; y_m > 0, \tag{7.3}$$

i.e. it does not have nonnegative solutions. Unrestricted variables so that this procedure would approach us a boundary of ODR, but it did not recede from it, i.e., so that the number of negative absolute

terms with each step/pitch would decrease, or, if the number of negative absolute terms it remains previous, then, at least, decreased their absolute values.

There is a series of methods of the selection of the solving cell/element for approach/approximation to the supporting/reference solution. Let us pause (without strict proof) at one of them.

Let there be one of equations (7.1) with negative absolute term. We seek in this row negative cell/element α_{ij}. If there is no this cell/element (all cell/elements $\alpha_{ij} > 0$), this is the sign/criterion of the fact that the system of equations (7.1) is incompatible with inequalities (7.3). It is real/actual, in the absence of negative cell/elements in row, entire/all right side of the corresponding equation can be only negative, and this contradicts the conditions of the nonnegative character of variables.

Let us assume that the negative cell/element is. Then is selected the column, in which it is located that as that solve.

Now it is necessary to select this column most solving cell/element. Let us consider all cell/elements of this column, which have identical sign with absolute term. From them let us select as that solving that, for which the relation to it of absolute term is

minimal.

Thus, is chosen the solving column, which solves cell/element in it and, which means, that the solving row.

We will be convinced based on example, as is accomplished approach/approximation to the supporting/reference solution with this rule of the selection of the solving cell/element. Incidentally we will be convinced of the soundness of this rule.

Page 73.

Example 1. To find (if it exists) the supporting/reference solution of the problem of linear programming with limitation-equalities:

$$\left.\begin{aligned} y_1 &= 1-(-x_1-2x_2+x_3), \\ y_2 &= -5-(-2x_1+x_2-x_3), \\ y_3 &= 2-(x_1+x_2), \\ y_4 &= 1-(-x_2+x_3), \end{aligned}\right\} \quad (7.4)$$

(here it is not brought the linear form which must be minimized, because the supporting/reference solution is ought irrespectively of the form of this form).

Solution. We record/write conditions (7.4) in the form of the standard table (see Table 7.1).

In Table 7.1 is negative absolute term by -5 in row y_2 of column x_1. According to rule, is selected any negative cell/element of this row, for example -2 (in table 7.1 it is emphasized). By this we selected the solving column x_1. As "candidates" to the role of the solving cell/element let us examine all those cell/elements of this column, which are differing by sign to their absolute term; this will be -2 and 1 (zero as the solving cell/element figure it cannot).

We compute for each of the "candidates" the relation to it of the absolute term:

$$(-5)/(-2) = 5/2; \quad 2/1=2.$$

Small from these relations to 2; that means cell/element 1 is selected as that solve and we transpose $x_1 \leftrightarrow y_3$ (see Table 7.2).

After the execution of actions, we come to taple 7.3.

In Table 7.3 as before one negative absolute term, but in absolute value it is already less than in table 7.1 - that means that we approach ODR.

Let us try to get rid also of this term. In row y_2 is only one negative cell/element -1 (it is emphasized). That means that the solving column can be only column x_3. We compute for all

cell/elements of this chair, which have identical sign with our **absolute term,** the ratio of absolute term to the cell/element:

$$3/1=3;\ (-1)/(-1)=1;\ 1/1=1.$$

Tables 7.1.

	Свободный член (1)	x_1	x_2	x_3
y_1	1	-1	-2	1
y_2	-5	$\underline{\underline{-2}}$	1	-1
y_3	2	1	1	0
y_4	1	0	-1	1

Key: (1). Absolute term.

Page 74.

Relation reaches the minimum, equal to 1, for two cell/elements; let us take and the quality of that solving the first of them (-1), that stands in row y_2 and column x_3, v let us do a replacement (see Table 7.4 and 7.5).

In Table 7.5 all absolute terms are nonnegative, and the supporting/reference solution is found:

$$y_3 = x_2 = y_2 = 0;\ y_1 = 2;\ x_3 = 1;\ x_1 = 2;\ y_4 = 0.$$

Example of 2. To find (if it exists) the supporting/reference solution of the system

$$\left.\begin{aligned} y_1 &= -4-(-x_1+2x_2), \\ y_2 &= -3-(x_1-x_2+x_3), \\ y_3 &= -10-(2x_1-x_2+x_3), \\ y_4 &= -2-(-x_1+x_2). \end{aligned}\right\} \quad (7.5)$$

Tables 7.2.

		Свободный член (1)	x_1 ($y_3 \leftrightarrow$)	x_2	x_3
	y_1	1 / 2	-1 / 1	-2 / 1	1 / 0
	y_2	-5 / 4	-2 / 2	1 / 2	-1 / 0
$x_1 \leftrightarrow$	y_3	2 / 2	1 / 1	1 / 1	0 / 0
	y_4	1 / 0	0 / 0	-1 / 0	1 / 0

Key: (1). Absolute term.

Tables 7.3.

	Свободный член (1)	y_3	x_2	x_3
y_1	3	1	-1	-1
y_2	-1	2	3	-1
x_1	2	1	1	0
y_4	1	0	-1	1

Key: (1). Absolute term.

Page 75.

Tables 7.4.

					$y_2 \leftrightarrow$
		Свободный член (1)	y_3	x_2	x_3
	y_1	3 / −1	1 / 2	−1 / 3	1 / $\boxed{1}$
$x_3 \leftrightarrow$	y_2	$\boxed{-1}$ / 1	$\boxed{2}$ / −2	$\boxed{3}$ / −3	−1 / −1
	x_1	2 / 0	1 / 0	1 / 0	0 / $\boxed{0}$
	y_4	1 / −1	0 / 2	−1 / 3	1 / $\boxed{1}$

Key: (1). Absolute term.

TABLE 7.5.

	Свободный член (1)	y_3	x_2	y_2
y_1	2	3	2	1
x_3	1	-2	-3	-1
x_1	2	1	1	0
y_4	0	2	2	1

Key: (1). Absolute term.

Tables 7.6.

	Свободный член (1)	x_1	x_2	x_3
y_1	-4	-1	2	0
y_2	-3	1	-1	1
y_3	-10	2	-1	1
y_4	-2	-1	1	0

Key: (1). Absolute term.

Page 76.

Solution. We record/write system of equations (7.5) in the form of the standard table (see Table 7.6).

It is selected row with the negative absolute term, for example, the first. In it there is negative cell/element (-1). Is selected column x_1 as that solve. We compute the relations:

$$(-4)/(-1)=4; \quad (-2)/(-1)=2.$$

Last/latter sense minimally; that means as that solve we take cell/element (-1) in row y_4 and we produce replacement $x_1 \leftrightarrow y_4$ (see Table 7.7 and 7.8).

Let us turn our attention to row y_3 in table 7.8. In it absolute term is negative, but there is not one negative cell/element (except quite absolute term). The corresponding equation takes the form:

$$y_3=-14-(2y_4+x_2+x_3).$$

Can with any nonnegative values y_4, x_2, x_3 value y_3 be nonnegative? It is obvious, no: with $y_4=x_2=x_3=0$ we will obtain $y_3=-14$, but increase y_4, x_2, x_3 over zero will do y_3 still less.

TABLE 7.7.

	Свободный член (1)	$y_4 \leftarrow$ x_1	x_2	x_3
y_1	-4 2	-1 -1	2 -1	0 0
y_2	-3 -2	1 1	-1 1	1 0
y_3	-10 -4	2 2	-1 2	1 0
$x_1 \leftarrow$ y_4	-2 2	-1 -1	1 -1	0 0

Key: (1). Absolute term.

Tables 7.8.

	Свободный член (1)	y_4	x_2	x_3
y_1	-2	-1	1	0
y_2	-5	1	0	1
y_3	-14	2	1	1
x_1	2	-1	-1	0

Key: (1). Absolute term.

Page 77.

Consequently, system (7.5) is incompatible with the inequalities, which ensue from the nonnegative character of variables, and the problem of linear programming with boundary conditions (7.5) the permissible solutions does not have. About the same testifies row y_2 table 7.8, where also there is not one negative cell/element (except quite absolute term).

Thus, we see that there is no need to specially trace the system of conditions of OZLP for consistency for the domain of the nonnegative solutions: this question is clarified automatically, in the process of the determination of supporting/reference solution.

8. Finding the optimum solution of the basic problem of linear programming.

In the previous paragraph we learned to find out the supporting/reference solution of system of equations of OZLP; during the searches of this supporting/reference solution, we were not completely occupied by the minimized function L. Now we will be occupied the optimization of solution, i.e., by finding such supporting/reference solution which converts into the minimum the linear function:

$$L = c_0 - (\gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_n x_n).$$

In §5 we already demonstrated the fundamental side of the methodology of the optimization of solution. Here we based on examples will show how this optimization can be carried out with the help of the tabular algorithm of replacement $x_j \leftrightarrow y_i$.

Example 1. To find the solution of the problem of linear programming with the equations

$$\left.\begin{aligned} y_1 &= 2-(x_1+x_2-2x_3),\\ y_2 &= 1-(x_1-x_2+x_3),\\ y_3 &= 5-(x_2+x_3),\\ y_4 &= 2-(2x_1-x_2), \end{aligned}\right\} \quad (8.1)$$

rotating in the minimum the linear function

$$L=0-(-x_1+2x_2+x_3). \quad (8.2)$$

Solution. All absolute terms in (8.1) are nonnegative, which means, that supporting/reference solution is present:

$$x_1=x_2=x_3=0;\quad y_1=2;\quad y_2=1;\quad y_3=5;\quad y_4=2.$$

Is it optimum? No, since the coefficients of x_2 and x_3 in (8.2) are positive, which means, increasing these variables, we reduce L.

Let us register (8.1) and (8.2) in the form of standard table (table 8.1).

Since the coefficients in the first row of x_2 and x_3 are positive, any of these variables can be deduced from the number of

free. Let this will be x_3. Which of the cell/elements of chair x_3 to take that solve? This cell/element must be positive. That means that of us exists the selection: 1 in row y_2 or 1 in row y_3. Let us select that them them, for which the relation to it of absolute term is minimal (proof see in §5).

Relations are equal to 1/1 = 1; 5/1=5. Minimum of them 1. That means that as that solve it is necessary to take cell/element 1 in chair x_3, row y_2. Let us replace $x_3 \leftrightarrow y_2$ (see Table 8.2, 8.3).

Page 78.

Tables 8.1.

	Свободный член (1)	x_1	x_2	x_3
L	0	-1	2	1
y_1	2	1	1	-2
y_2	1	1	-1	1
y_3	5	0	1	1
y_4	2	2	-1	0

Key: (1). Absolute term.

TABLE 8.2.

	Свободный член (1)	x_1	x_2	$y_2 \leftrightarrow$ x_3
L	0 -1	-1 -1	2 1	1 $\boxed{-1}$
y_1	2 2	1 2	1 -2	-2 $\boxed{2}$
$x_3 \leftrightarrow$ y_2	$\boxed{1}$ 1	$\boxed{1}$ 1	$\boxed{-1}$ -1	1 1
y_3	5 -1	0 -1	1 1	1 $\boxed{-1}$
y_4	2 0	2 0	-1 0	0 $\boxed{0}$

Key: (1). Absolute term.

Tables 8.3.

	Свободный член (1)	x_1	x_2	y_2
L	-1	-2	3	-1
y_1	4	3	-1	2
x_3	1	1	-1	1
y_3	4	-1	2	-1
y_4	2	2	-1	0

Key: (1). Absolute term.

Page 79.

In upper row table 8.3 is the positive coefficient of x_2, which means, that x_2 it is necessary to deduce from unrestricted variables. Is selected as the solving that positive cell/element chair x_2, for which the relation to it of absolute term is minimal. But in chair x_2 unique positive cell/element 2, it is selected as that solve (see Table 8.4 and 8.5).

It turns out that procedure is not still finished: in the first row table 8.5, is a positive cell/element in column y_2, which means, that the variable y_2 must be deduced from the number of free. As that solve we take that of the positive cell/elements of chair y_2, for which the relation to it of absolute term is minimal. Equate/comparing the relaticns

$$6 : {}^3/_2 = 4, \quad 3 : {}^1/_2 = 6,$$

it is selected as that solving cell/element 3/2 in row y_1 and chair y_2 we continue the procedure of the optimization (see Table 8.6 and 8.7).

In the first row table 8.7 there is not one positive cell/element; that means optimum solution reached; it will be:

$$x_1 = y_3 = y_1 = 0; \quad y_2 = 4; \quad x_3 = 1; \quad x_2 = 4; \quad y_4 = 6.$$

At these values of variables, the linear function L reaches its

minimum value equal to

$$L_{min} = -9.$$

Does arise the question: a that if in the chair, which contains positive row element L, will be located not one positive cell/element, in order to make it solving? It is easy to ascertain that in this case function L is not limited from below and OZLP does not have optimum solution.

It is real/actual, in this case an increase of the variable, that corresponds to this chair, reduces the linear function L and cannot do one of the base variables negative, which means, that nothing impedes the unlimited decrease of function L.

Thus, let us formulate the rules of the determination of the optimum solution of OZLP by the simplex method.

1. If all absolute terms (without considering row L) in simplex-table are nonnegative, but in row L (without considering absolute term) there is not one positive cell/element, then optimum solution is reached.

TABLE 8.4.

	Свободный член	x_1	x_2 ($y_3 \leftrightarrow$)	y_2
L	-1 / -6	-2 / $\frac{3}{2}$	3 / $-\frac{3}{2}$	-1 / $\frac{3}{2}$
y_1	4 / 2	3 / $-\frac{1}{2}$	-1 / $\frac{1}{2}$	2 / $-\frac{1}{2}$
x_3	1 / 2	1 / $-\frac{1}{2}$	-1 / $\frac{1}{2}$	1 / $-\frac{1}{2}$
$x_2 \leftrightarrow$ y_3	4 / 2	-1 / $-\frac{1}{2}$	2 / $\frac{1}{2}$	-1 / $-\frac{1}{2}$
y_4	2 / 2	2 / $-\frac{1}{2}$	-1 / $\frac{1}{2}$	0 / $-\frac{1}{2}$

Key: (1). Absolute term.

Page 80.

Tables 8.5.

	Свободный член (1)	x_1	y_3	y_2
L	-7	$-\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$
y_1	6	$\frac{5}{2}$	$\frac{1}{2}$	$\frac{3}{2}$
x_3	3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
x_2	2	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$
y_4	4	$\frac{3}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$

Key: (1). Absolute term.

Tables 8.6.

$y_1 \leftrightarrow$

	Свободный член (1)	x_1	y_3	y_2
L	-7 $\quad -2$	$-\frac{1}{2}$ $\quad -\frac{5}{6}$	$-\frac{3}{2}$ $\quad -\frac{1}{6}$	$\frac{1}{2}$ $\quad \boxed{-\frac{1}{3}}$
$y_2 \leftrightarrow$ y_1	$\boxed{6}$ $\quad 4$	$\boxed{\frac{5}{2}}$ $\quad \frac{5}{3}$	$\boxed{\frac{1}{2}}$ $\quad \frac{1}{3}$	$\frac{3}{2}$ $\quad \frac{2}{3}$
x_3	3 $\quad -2$	$\frac{1}{2}$ $\quad -\frac{5}{6}$	$\frac{1}{2}$ $\quad -\frac{1}{6}$	$\frac{1}{2}$ $\quad \boxed{-\frac{1}{3}}$
x_2	2 $\quad 2$	$-\frac{1}{2}$ $\quad \frac{5}{6}$	$\frac{1}{2}$ $\quad \frac{1}{6}$	$-\frac{1}{2}$ $\quad \boxed{\frac{1}{3}}$
y_4	4 $\quad 2$	$\frac{3}{2}$ $\quad \frac{5}{6}$	$\frac{1}{2}$ $\quad \frac{1}{6}$	$-\frac{1}{2}$ $\quad \boxed{\frac{1}{3}}$

Key: (1). Absolute term.

Tables 8.7.

	Свободный член (1)	x_1	y_3	y_1
L	-9	$-\frac{4}{3}$	$-\frac{5}{3}$	$-\frac{1}{3}$
y_2	4	$\frac{5}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
x_3	1	$-\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$
x_2	4	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
y_4	6	$\frac{7}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

Key: (1). Absolute term.

Page 81.

Tables 8.8.

	Свободный член (1)	x_1	x_2	x_3	x_4
L	0	-2	1	0	0
y_1	0	-1	1	0	0
y_2	2	0	1	-1	0
y_3	1	0	0	-1	-1

Key: (1). Absolute term.

Tables 8.9.

	Свободный член (1)	x_1	$y_1 \leftrightarrow$ x_2	x_3	x_4
L	0 / 0	-2 / 1	1 / $\boxed{-1}$	0 / 0	0 / 0
$x_1 \leftrightarrow$ y_1	$\boxed{0}$ / 0	$\boxed{-1}$ / -1	1 / 1	$\boxed{0}$ / 0	$\boxed{0}$
y_2	2 / 0	0 / 1	1 / $\boxed{-1}$	-1 / 0	0 / 0
y_3	1 / 0	0 / 0	0 / $\boxed{0}$	-1 / 0	-1 / 0

Key: (1). Absolute term.

TABLE 8.10

	Свободный член (1)	x_1	y_1	x_3	x_4
L	0	-1	-1	0	0
x_2	0	-1	1	0	0
y_2	2	1	-1	-1	0
y_3	1	0	0	-1	-1

Key: (1). Absolute term.

Page 82.

2. If in row L there is positive cell/element, but in column, which corresponds to it, there is not one positive cell/element, then linear function L is not limited from below, and optimum solution does not exist.

3. If in this chair there are positive cell/elements, then one should replace of one of unrestricted variables by one of basis, moreover as that solve it is necessary to take that cell/element of this chair, for which relation to it of corresponding absolute term is minimal.

In conclusion let us pause at the so-called "degenerate" case when one (or more) absolute terms in equation-limitations it is

obtained equal to zero. This means that in this supporting/reference solution are converted into zero only unrestricted variables, but also some of the base. Let us consider an example.

Example of 2. To find the solution of the problem of linear programming with the conditions

$$\left.\begin{array}{l} y_1 = x_1 - x_2, \\ y_2 = -x_2 + x_3 + 2, \\ y_3 = x_3 + x_4 + 1, \end{array}\right\} \qquad (8.3)$$

rotating in the minimum the linear function

$$L = 2x_1 - x_2. \qquad (8.4)$$

Solution. We record/write (8.3) and (8.4) in the form of the standard table (see Table 8.8).

According to general rule, we seek in chair x_2 the solving cell/element, for which the relation to it of absolute term is nonnegative and it is minimal. Equating relation to 0:1 and 2:1, we are stopped on solving cell/element 1 in row y_1, for which this sense is equal to zero. We produce replacement $x_2 \longleftrightarrow y_1$ (see Table 8.9 and 8.10).

During transition from one table 8.8 to next 8.10, it is logical, did not occur the decrease of the linear function L (it both

was and it remained equal tc zero), but the cell/elements of upper row they became everything nonpositive, from what evident that the optimum solution is reached: the minimum of function was equal to zero and is reached at $x_1=y_1=x_3=x_4=0$; $x_2=0$; $y_2=2$; $y_3=1$.

Let us do still one, the latter, observaticn apropos of the so-called "ringing". We already saw that in the presence of "degeneration" it can seem that replacement of cne of the unrestricted variables by base and back leads only to the exchange of variables, without the decrease of the linear function L. In very rare cases can seem that the consecutive application/use of a rule of the selection of the solving cell/element leads to the fact that after several replacements $x_j \leftrightarrow y_i$ we again are returned to the same set of base and unrestricted variables, from which they began. This is called "ringing". Virtually in order to avoid this, sufficiently it is during repetition to take the solving cell/element not in the manner that it was undertaken for the first time (for example, in other chair). During the organization of the algorithm of linear programming by ETsVM into program, must be intrcduced corresponding indication.

Page 83.

9. Transport problem of linear programming.

In the previous paragraphs the simplex method of the solution of problem of linear programming presented is universal and it is applicable for the solution of any such problems. However, there are some particular types of problems of the linear programming which, by the force of some special feature/peculiarities of its structure, admit solution by simpler methods. To it is related, in particular, the so-called transport problem.

The classical transport problem of linear programming is formulated as follows.

There is m of point/items of the sending: $A_1, A_2, \ldots, A_m$, in which are concentrated the supplies of some uniform goods (load) in a quantity respectively $a_1, a_2, \ldots, a_m$ of unity. Furthermore, is n of stations of destination: $B_1, B_2, \ldots, B_n$, feeding claims respectively to $b_1, b_2, \ldots, b_n$ unity of goods.

It is assumed that the sum of all claims is equal to the sum of all supplies:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{9.1}$$

Is known the cost/value c_{ij} of the transport of unity of goods

from each point/item of sending A_i to each station of destination B_j. The table (matrix/die) of the cost/values of transport c_{ij} is assigned:

$$\begin{Vmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{Vmatrix}.$$

It is required to comprise such plan/layout of transport, by which all claims would be carried out, and in this case the common/general/total cost/value of all transport was minimum.

Upon this formulation of the problem the index of the efficiency of the plan/layout of transport is the cost/value; therefore stated problem more precisely calls transport problem in the criterion of cost/value.

Let us give to this problem mathematical formulation. Let us designate x_{ij} - quantity of load, transmitted from the i point/item of sending A_i for j- station of destination B_j (i=1, ..., j=1, ..., n). Nonnegative variables $x_{11}, x_{12}, \dots, x_{mn}$ (number of which, obviously, equally mxn) must satisfy the following conditions:

1. The total quantity of load, directed from each point/item of sending in all the stations of destination, must be equal to the supply of the load in this point/item.

Page 84.

This will give to us m condition-equalities:

$$\left.\begin{array}{l} x_{11}+x_{12}+\ldots+x_{1n}=a_1, \\ x_{21}+x_{22}+\ldots+x_{2n}=a_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{m1}+x_{m2}+\ldots+x_{mn}=a_m, \end{array}\right\}$$

or, it is shorter,

$$\left.\begin{array}{l} \sum_{j=1}^{n} x_{1j}=a_1, \\ \sum_{j=1}^{n} x_{2j}=a_2, \\ \ldots\ldots\ldots \\ \sum_{j=1}^{n} x_{mj}=a_m. \end{array}\right\} \qquad (9.2)$$

2. Total quantity of load, supply/delivered to each station of destination from all point/items of sending, must be equal to claim, subject this point/item. This will give n condition-equalities:

$$\left.\begin{array}{l} x_{11}+x_{21}+\ldots+x_{m1}=b_1, \\ x_{12}+x_{22}+\ldots+x_{m2}=b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{1n}+x_{2n}+\ldots+x_{mn}=b_n, \end{array}\right\}$$

or, it is shorter,

$$\left.\begin{array}{l} \sum_{i=1}^{m} x_{i1}=b_1, \\ \sum_{i=1}^{m} x_{i2}=b_2, \\ \ldots\ldots\ldots \\ \sum_{i=1}^{m} x_{in}=b_n. \end{array}\right\} \qquad (9.3)$$

3. Total cost/value of all transport, i.e., sum of values x_{ij}, multiplied by appropriate cost/values c_{ij}, must be minimum:

$$L = c_{11}x_{11} + c_{12}x_{12} + \dots + c_{1n}x_{1n} + \\ + c_{21}x_{21} + c_{22}x_{22} + \dots + c_{2n}x_{2n} + \dots + \\ + c_{m1}x_{m1} + c_{m2}x_{m2} + \dots + c_{mn}x_{mn} = \min,$$

or, much shorter,

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}x_{ij} = \min, \quad (9.4)$$

where sign of double sum $\sum_{i=1}^{m} \sum_{j=1}^{n}$ means that addition is produced on all combinations of indices (i=1, ..., m; j=1, ..., n), i.e., on all combinations of point/items of sending with stations of destination.

Page 85.

Function (9.4) is linear, limitation - equality (9.2), (9.3) are also linear. Before us - the typical problem of linear programming with limitation-equalities (OZLP).

Like any other problem of linear programming, it it would be possible to solve by the simplex method, but this problem has some special feature/peculiarities, which make it possible to solve it a more simply. Reason is the fact that all coefficients of variables in equations (9.2), (9.3) are equal to one. Furthermore, has a value the structure of communication/connections between conditions. It is not

difficult to ascertain that not all m+n equations of our problem are independent variables. It is real/actual, store/adding up between themselves all the equations (9.2) and all the equations (9.3), we must obtain one and the same, by the force of condition (9.1). Thus, conditions (9.2), (9.3) are connected by one linear dependence, and actually of these equations only m + n - 1, but not m + n are the linearly independent. That means that the rank of system of equations (9.2), (9.3) is equal to

$$r = m + n - 1,$$

and therefore, it is possible to solve these equations relative to m + n - 1 base variables, after expressing them through the others, free.

Let us count a quantity of unrestricted variables. It is equal to:

$$k = mn - (m + n - 1) = mn - m - (n - 1) = \\ = m(n - 1) - (n - 1) = (m - 1)(n - 1).$$

We know that in the problem of linear programming the optimum solution is achieved in one of the apex/vertexes of ODR where at least k of variables are converted into zero. That means that in our case for the optimum plan/layout of the transport at least (m - 1) (n - 1) of values x_{ij} they must be equal to zero.

Let us agree the terminology. The values x_{ij} of a quantity of

ones of load, directed from point/item A_i in point/item B_j we will call transport.

Any value part (x_{ij}) (i=1, ..., m; j=1, ..., n) let us call the plan/layout of transport, or it is simple by plan/layout.

Plan/layout (x_{ij}) let us call permissible, if it satisfies conditions (9.2), (9.3) (the so-called "balance conditions"): all claims are satisfied, all supplies exhausted.

The permissible plan/layout let us call supporting/reference, if in it are different from zero not more than r=m+n-1 base transport x_{ij}, and remaining transport are equal to zero.

Plan/layout (x_{ij}) let us call optimum, if it, among all permissible plan/layouts, leads to the smallest cost/value of all transport.

Page 86.

Let us pass to the presentation of the methods of the solution of transport problem (TZ). These methods do not require manipulations with simplex-tables, but they are reduced to simpler operations directly with the table where in the determined order are registered

all conditions of TZ. This table we will call transport table.

In transport table are record/written

- the point/items of sending and designaticn/purposes,

- the supplies, availatle in the pcint/items of sending,

- the claims, subject by stations cf destination,

- the cost/value of transport from each pcint/item of sending into each station of destination.

The cost/values of transport we will place in the upper right-hand corner of each nucleus, with the fact in order in nucleus itself with compilation of plan/layout to place the transport x_{ij}.

The specimen/sample table gives in table 9.1.

For brevity subsequently, let us designate the point/items of sending - PO, stations of destinaticn - PN. In the upper right-hand corner of each cage/cell wrcte the cost/values of the transport of cne of goods (load) from PO A_i into PN B_j. In right column placed the supplies of goods in each PO, in lcwer row - the claims, the

subjects by each PN. For TZ the sum of supplies is equal to the sum of claims; the common/general/total value of this sum is record/written in the right lower nucleus of table.

Above we showed that the rank of the system of equation-limitations of TZ was equal to r=m+n-1, where m - a number of rows, and n - a number of columns of transport table. That means that in each supporting/reference plan/layout, including optimum, they will be different from zero not more than n+m-1 transport.

The nuclei (cage/cell) of the tables in which we will record/write these different from zero transport, let us agree to call base, and the others (empty) with free.

TABLE 9.1.

ПН / ПО	B_1	B_2	......	B_n	Запасы a_i (1)
A_1	c_{11}	c_{12}	......	c_{1n}	a_1
A_2	c_{21}	c_{22}	......	c_{2n}	a_2
⋮	...	...	......	...	⋮
A_m	c_{m1}	c_{m2}	......	c_{mn}	a_m
Заявки b_j (2)	b_1	b_2	......	b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

Key: (1). Supplies. (2). Claims.

Page 87.

Thus, the solution of TZ was reduced to the following. To find such values of the positive transport which, being are written in elementary cells of transport table, would satisfy the following conditions:

- sum of transport of each table row must be equal to reserve of the given PO;

- sum of transport of each chair must be equal to the claim of

given PN;

- common/general/total cost/value of transport - minimum.

In the future all the actions on the determination of the solution of TZ will be reduced to the transformation of transport table 9.1.

During the description of these transformations to us it is convenient it will be to use the numbering of the cage/cells of table (similar numbering of the cage/cells of the chessboard). By cage/cell (A_i, B_j) or, it is shorter, cage/cell (i, j) we will call the cage/cell, which stands in the i row and the j chair of transport table. For example, uppermost left cage/cell will be designated (1, 1), that stands hearth by it (2, 1) and so forth.

10. Determination of supporting/reference plan/layout.

The solution of transport problem as any problem of linear programming, begins from the determination of supporting/reference solution, or as we will speak, supporting/reference plan/layout. Unlike the general case of CZLP with arbitrary limitations and the minimized function, the solution of TZ always exists. It is real/actual, from purely physical considerations it is clear that

although some permissible plan/layout to exist must among the permissible plan/layouts without fail is optimum (it can be, not one), because the linear function L - a cost/value of transport is knowingly nonnegative (it is limited from below by zero). In this paragraph we will show how to construct supporting/reference plan/layout. For this, there are different methods from which we will pause at simplest, the so-called "method of the northwest corner". To clarify it most simple will be based on specific example.

Example 1. Conditions of TZ are assigned by the transport table (see Table 10.1).

It is required to find the supporting/reference solution of TZ (to construct supporting/reference plan/layout).

Solution. Let us rewrite table 10.1 and will fill it with transport gradually, beginning with left upper nucleus (1.1) ("northwest corner" of table). Let us discuss in this case as follows. Point/item B_1 fed claim for 18 ones of load. Let us satisfy this claim of volt-ampere the calculation of supply 48, available in point/item A_1, and let us register transport by 18 in cage/cell (1.1). After this claim of point/item B_1 it is satisfied, and in point/item A_1, remained an additional 30 ones of load. Let us satisfy because of them the claim of point/item B_2 (27 ones), let us register

AD-A066 460 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/2

OPERATIONS RESEARCH. PART I.(U)

JUN 78 Y S VENTTSEL'

UNCLASSIFIED FTD-ID(RS)T-0687-78-PT-1 NL

3 OF 6

AD A066460

27 in cage (1.2); the remaining 3 units of point/item A_1 let us assign to point/item B_3. In the composition of the claim of point/item B_3 remained not satisfied of 39 unity.

Page 88.

From them 30 by cut because of point/item A_2, than its supply it will be exhausted, and an additional 9 let us take from point/item A_3. From the remaining 18 units of point/item A_3 12, let us isolate to point/item B_4; remaining 6 units let us assign to point/item B_5, which together with all 20 units of point/item A_4 will cover its claim (see Table 10.2).

On this, safety distribution it is finished: each station of destination obtained load according to its claim. This is expressed in the fact that the sum of transport of each row is equal to the appropriate supply, and in chair - claim.

Thus, by us immediately is comprised the plan/layout of transport, which satisfies balance conditions. The obtained solution is only not permissible, but also supporting/reference the solution of transport problem.

TABLE 10.1.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i
A_1	^{10}	^{8}	^{5}	^{6}	^{9}	48
A_2	^{6}	^{7}	^{8}	^{6}	^{5}	30
A_3	^{8}	^{7}	^{10}	^{8}	^{7}	27
A_4	^{7}	^{5}	^{4}	^{6}	^{8}	20
Заявки b_j (2)	18	27	42	12	26	125

Key: (1). Supplies. (2). Claims.

Tables 10. 2.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы a_i (1)
A_1	18 ^{10}	27 ^{8}	3 ^{5}	^{6}	^{9}	48
A_2	^{6}	^{7}	30 ^{8}	^{6}	^{5}	30
A_3	^{8}	^{7}	9 ^{10}	12 ^{8}	6 ^{7}	27
A_4	^{7}	^{5}	^{4}	^{6}	20 ^{8}	20
Заявки b_j (2)	18	27	42	12	26	125

Key: (1). Supplies. (2). Claims.

Page 89.

The cage/cells of the tables in which stand nonzero transport, are base, their number satisfies condition $r=m+n-1=8$. The others cells - free (empty), in them stand nonzero transport, their number is equal $(n-1)(m-1)=12$. That means that our plan/layout-supporting/reference and stated problem of the construction of supporting/reference plan/layout is solved.

Does arise the question: a is this plan/layout optimum for cost/value? It goes without saying that no ! Indeed with its construction we in no way considered the cost/values of transport c_{ij}. It is logical, plan/laycut was not obtained optimum. It is real/actual, the cost/value of this plan/layout which will be located, if we multiply each transport by the appropriate cost/value, it is equal to $18\cdot 10+27\cdot 8+3\cdot 5+30\cdot 8+9\cdot 10+12\cdot 8+6\cdot 7+20\cdot 8=1039$.

To improve this plan/layout, after transferring, for example, 18 units from cage/cell (1.1) intc cell (2.1) and, in order not to break balance, after transferring the same of 18 units from cage/cell (2.3) into cell (1.3). We will obtain the new plan/layout, given in table 10.3.

It is not difficult to ascertain that the cost/value of new plan/layout is equal to $27\cdot 8+21\cdot 5+18\cdot 6+12\cdot 8+9\cdot 10+12\cdot 8+6\cdot 7+20\cdot 8=913$, i.e., per 126 units is smaller than the cost/value of the

plan/layout, given in table 10.3.

Thus, because of the cyclic permutaticn of 18 units of the load of some cage/cells of cthers we succeeded in reducing the cost/value of plan/layout. On this method of decreasing the cost/value subsequently will be based the algorithm of the optimization of the plan/layout of transport.

Let us pause at one special feature/peculiarity of the plan/layout of transport, which can be met both during the construction of supporting/reference plan/layout and during its improvement. Speech occurs about the so-called "degenerate" plan/layout in which some of the base transport prove to be equal to zero. Let us consider a specific example of the emergence of the degenerate plan/layout.

Example of 2. Is given transport table (without the cost/values of transport, since we are dealing only with the construction of supporting/reference plan/layout) - see Table 10.4.

TABLE 10.3

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы a_i (1)
A_1	(10)	27 (8)	21 (5)	(6)	(9)	48
A_2	18 (6)	(7)	12 (8)	(6)	(5)	30
A_3	(8)	(7)	9 (10)	12 (8)	6 (7)	27
A_4	(7)	(5)	(4)	(6)	20 (8)	20
Заявки b_j (2)	18	27	42	12	26	125

Key: (1). Supplies. (2). Claims.

Page 90.

Tables 10.4.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы a_i (1)
A_1						20
A_2						30
A_3						25
A_4						20
Заявки b_j (2)	10	10	20	35	20	95

Key: (1). Supplies. (2). Claims.

Tables 10.5.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы a_i (1)
A_1	10	10				20
A_2			20	10		30
A_3				25		25
A_4					20	20
Заявки b_j (2)	10	10	20	35	20	95

Key: (1). Supplies. (2). Claims.

Tables 10.6.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы a_i (1)
A_1	10	10	ε			$20+\varepsilon$
A_2			$20-\varepsilon$	$10+\varepsilon$		30
A_3				$25-\varepsilon$	2ε	$25+\varepsilon$
A_4					$20-2\varepsilon$	$20-2\varepsilon$
Заявки b_j (2)	10	10	20	35	20	95

Key: (1). Supplies. (2). Claims.

Page 91.

To comprise the supporting/reference plan/layout of transport.

Solution. Applying the method of the northwest corner we will obtain table 10.5.

Supporting/reference plan/layout is comprised. Its special feature/peculiarity is the fact that in it only six, but not eight different from zero transport. This means, certain of the base transport which must be m+n-1=8, they render/showed equal to zero.

It is not difficult to note that why this occurred: during the distribution of supplies according to stations of destination in certain cases, the residue/remainders proved to be equal to zero and into the appropriate cage/cell did not fall.

Such cases of "degeneration" can appear not only during the composition of supporting/reference plan/layout, but also during its transformation, optimization.

In the future to us is convenient will be always to have in the transport table m+n-1 of elementary cells, although in some of them perhaps they will stand the zero values of transport. For this, it is possible to negligibly little change supplies or claims, so as to total balance would not be broken, but excess, "intermediate"

balances were destroyed. Is sufficient in necessary places to change supplies or the claims, for example, to value ε, a after the determination of the optimum solution to assume $\varepsilon = 0$.

Let us show how to pass from the degenerate plan/layout to that nondegenerate based on the example table 10.5. Let us change slightly supplies in the first row and will place them equal to $20+\varepsilon$. Furthermore, in the third row let us write supplies $25+\varepsilon$. In order "to reduce balance", in the fourth row we place supplies $20-2\varepsilon$ (see Table 10.6). For this table we construct supporting/reference plan/layout by the method of the northwest corner.

In Table 10.6 has already been contained as many base variables, as is required: $m+n-1=8$. In the future, after the optimization of plan/layout, it will be possible to assume $\varepsilon = 0$.

11. Improvement in the plan/layout of transport. Cycle of recalculation.

In the previous paragraph we already rapidly were introduced to the method of an improvement in the plan/layout, consisting of the fact that some transport, without the damage of balance, are transferred from cage/cell to cage/cell on certain closed cycle. Here we will consider these cyclic permutations in more detail.

Let us take the transport table, which consists, for example, of m=5 rows and n=6 chairs (number of rows and columns is unessential).

Cycle in transport table we will call several cage/cells, connected locked broken line which in each cage/cell accomplishes rotation on 90°.

For example, table 11.1 depicts two cycles: the first with four apex/vertexes (2.1), (2.3), (4.3), (4.1) and the second - with eight apex/vertexes (1.4), (1.6), (4.6), (4.4), (3.4), (3.5), (5.5), (5.4). By rifleman/pointers is shown the direction of the circuit/bypass of cycle.

Page 92.

TABLE 11.1.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	B_6	Запасы a_i (1)
A_1	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	a_1
A_2	c_{21}	c_{22}	c_{23}	c_{24}	c_{25}	c_{26}	a_2
A_3	c_{31}	c_{32}	c_{33}	c_{34}	c_{35}	c_{36}	a_3
A_4	c_{41}	c_{42}	c_{43}	c_{44}	c_{45}	c_{46}	a_4
A_5	c_{51}	c_{52}	c_{53}	c_{54}	c_{55}	c_{56}	a_5
Заявки b_j (2)	b_1	b_2	b_3	b_4	b_5	b_6	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

Key: (1). Supplies. (2). Claims.

Tables 11.2.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	B_6	Запасы a_i (1)
A_1	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	a_1
A_2	c_{21}	c_{22}	c_{23}	c_{24}	c_{25}	c_{26}	a_2
A_3	c_{31}	c_{32}	c_{33}	c_{34}	c_{35}	c_{36}	a_3
A_4	c_{41}	c_{42}	c_{43}	c_{44}	c_{45}	c_{46}	a_4
A_5	c_{51}	c_{52}	c_{53}	c_{54}	c_{55}	c_{56}	a_5
Заявки b_j (2)	b_1	b_2	b_3	b_4	b_5	b_6	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

Key: (1). Supplies. (2). Claims.

Page 93.

It is not difficult to ascertain that each cycle has even number cf apex/vertexes and, which means, that even number of component/links (arrow/pointers).

Let us agree to note by sign "+" those apex/vertexes of cycle, in which the transport increase, and by sign "-" - those apex/vertexes in which they are reduced. Cycle with the noted apex/vertexes let us call "designated". Table 11.2 shows two designated cycles: first Ts_1 with four apex/vertexes (1.1), (1.2), (3.2) and (3.1) and second Ts_2 with eight apex/vertexes (3.4), (3.6), (5.6), (5.3), (2.3), (2.5), (4.5) and (4.4).

To transfer (to "move") some quantity of units of load on the designated cycle - this means to increase the transport, which stand in the positive apex/vertexes of cycle, to this quantity of units, and transport, which stand in negative apex/vertexes - to decrease by the same quantity. It is obvious, during the transfer of any number of units on cycle, the equilibrium between supplies and claims does not vary: as before the sum of transport of each row is equal to the supplies of this row, but the sum of transport of each chair - claim

of this column. Thus, during any end-around carry, which leaves transport nonnegative, the permissible plan/layout it remains permissible. The cost/value of plan/layout in this case can vary - increase or be reduced.

Let us name the value of cycle an increase in the cost/value of transport during transferring of one unit of load over the designated cycle. It is obvious, the value of cycle is equal to the algebraic sum of the cost/values, which stand of the apex/vertexes of cycle, moreover the cost/values, which stand in positive apex/vertexes, are taken with sign "+", and in negative - with sign "-". For example, for a cycle Ts_1 in table 11.2 values is equal to:

$$c_{11}-c_{12}+c_{32}-c_{31},$$

while for a cycle Ts_2

$$c_{34}-c_{36}+c_{56}-c_{53}+c_{23}-c_{25}+c_{45}-c_{44}.$$

Let us designate the value of cycle Ts through γ. During transferring of one unit of load over cycle Ts, the cost/value of transport increases by value γ; during transferring over it k of the units of load the cost/value of transport increases by $k\gamma$.

It is obvious, for an improvement in the plan/layout, has the sense to move transport only on that cycles whose value is negative.

Each time when to us be managed to complete this transferring, cost of plan/layout is reduced by the appropriate value $k\gamma$.

Since transport cannot be negative, we will use only such cycles whose negative apex/vertexes lie/rest at elementary cells of the table where stand positive transport [1].

FOOTNOTE [1]. In the case of degeneration, as we will see further, can render/show useful fictitious transfer on the cycle whose negative apex/vertex lie/rests at cage/cell with zero transport. ENDFOOTNOTE.

If cycles with negative value in table no longer remained, this means that further improvement in the plan/layout is impossible, i.e., optimum plan/layout is achieve/reached.

Page 94.

The method of a consecutive improvement in the plan/layout of transport lies in the fact that in table are found out the cycles with negative value, on them are moved the transport, and plan/layout is improved until cycles with negative value no longer remain.

During an improvement in the plan/layout by end-around carries, as a rule, they use the method, borrowed from the simplex method: with each step/pitch (cycle) they substitute one unrestricted variable by base, i.e., is filled one free cage/cell and instead of that they free/release one of elementary cells. In this case, the total number of elementary cells remains by constant/invariable and equal $m + n-1$. This method is convenient in that for it is more easily found the adequate/approaching cycles.

It is possible to demonstrate that for any free cage/cell of transport table always there is a cycle (and besides only), one of apex/vertexes of which lie/rests at this free cage/cell, and all others - in elementary cells. If the value of this cycle, with plus

in free cage/cell, is negative, then plan/layout can be improved transferring by of transport over this cycle. A quantity of unity of load k which can be moved, is determined by the minimum value of the transport, which stand in the negative apex/vertexes of the cycle (if we move the larger number of unity of load, will arise negative transport).

Example 1. To find optimum plan/layout for the transport problem, given in table 11.3.

Solution. We comprise supporting/reference plan/layout by the method of the northwestern angle (table 11.4).

The cost/value of this plan/layout is equal to:

$$L_1 = 22\cdot 10 + 9\cdot 7 + 25\cdot 6 + 23\cdot 5 + 18\cdot 6 + 20\cdot 7 = 796.$$

The number of base variables, as it is set/assumed in the nondegenerate case, is equal to $r = m + n - 1 = 3 + 4 - 1 = 6$.

Let us try to improve plan/layout, after occupying free cage/cell (2.4) with minimum cost/value 4. Cycle, which corresponds to this cage/cell, shown in table to 11.4. The value of this cycle is equal to $\gamma = 4 - 7 + 6 - 5 = -2$.

On this cycle we can mcve the maximum of 20 unity of the load (in order not to obtain in the cage/cell (3.4) of negative transport). The new, improved plan/laycut is shcwn in table to 11.5.

Page 95.

Table 11.3.

ПН / ПО	B_1	B_2	B_3	B_4	Запасы (1) a_i
A_1	^{10}	^{7}	^{6}	^{8}	31
A_2	^{5}	^{6}	^{5}	^{4}	48
A_3	^{8}	^{7}	^{6}	^{7}	38
Заявки (2) b_j	22	34	41	20	117

Key: (1). Supplies. (2). Claims.

Table 11.4.

ПН / ПО	B_1	B_2	B_3	B_4	Запасы (1) a_i
A_1	22 ^{10}	9 ^{7}	^{6}	^{8}	31
A_2	^{5}	25 ^{6}	23 − ^{5}	+ ^{4}	48
A_3	^{8}	^{7}	18 + ^{6}	− 20 ^{7}	38
Заявки (2) b_j	22	34	41	20	117

Key: (1). Supplies. (2). Claims.

Table 11.5.

ПН / ПО	B_1	B_2	B_3	B_4	Запасы (1) a_i
A_1	22 − ^{10}	9 + ^{7}	^{6}	^{8}	31
A_2	+ ^{5}	− 25 ^{6}	3 ^{5}	20 ^{4}	48
A_3	^{8}	^{7}	38 ^{6}	^{7}	38
Заявки (2) b_j	22	34	41	20	117

Key: (1). Supplies. (2). Claims.

Table 11.6.

ПН / ПО	B_1	B_2	B_3	B_4	Запасы (1) a_i
A_1	^{10}	31 ^{7}	^{6}	^{8}	31
A_2	22 ^{5}	3 ^{6}	3 ^{5}	20 ^{4}	48
A_3	^{8}	^{7}	38 ^{6}	^{7}	38
Заявки (2) b_j	22	34	41	20	117

Key: (1). Supplies. (2). Claims.

Page 96.

Table 11.7.

ПН / ПО	B_1	B_2	B_3	Запасы (1) a_i
A_1	^{10}	^{5}	^{4}	40
A_2	^{6}	^{4}	^{5}	23
A_3	^{7}	^{3}	^{6}	20
Заявки (2) b_j	20	20	43	83

Key: (1). Supplies. (2). Claims.

Table 11.8.

ПН / ПО	B_1	B_2	B_3	Запасы (1) a_i
A_1	20 ^{10}	20 $-$ ^{5}	$+$ ε ^{4}	$40+\varepsilon$
A_2	^{6}	^{4}	23 ^{5}	23
A_3	^{7}	$+$ ^{3}	$-$ $20-\varepsilon$ ^{6}	$20-\varepsilon$
Заявки (2) b_j	20	20	43	83

Key: (1). Supplies. (2). Claims.

Table 11.9.

ПН / ПО	B_1	B_2	B_3	Запасы a_i
A_1	20 − [10]	ε [5]	+ 20 [4]	$40+\varepsilon$
A_2	+ [6]	[4]	23 − [5]	23
A_3	[7]	$20-\varepsilon$ [3]	[6]	$20-\varepsilon$
Заявки b_j	20	20	43	83

Key: (1). Supplies. (2). Claims.

Page 97.

Cost/value of this plan/layout $L_2 = 796 + 20\cdot(-2) = 756$. In it as previous are six elementary cells.

For further improvement in the plan/layout, let us focus attention on free cage/cell (2. 1) with cost/value 5. The cycle, which corresponds to this cage/cell, shown in Table 11.5; value its 7 - 6 + 5 - 10 = -4 on this cycle let us move 22 unity of load how we decrease the cost/value of transport to $L_3 = 756 + 22\cdot(-4) = 668$ (see Table 11.6).

Let us try to further improve this plan/layout, counting the values of cycles, that begin by positive apex/vertex in free cage/cell. we examine/scan the available free cage/cells Table 11.6 we determine the value of cycle for each of them. All these values (we let for reader to check this) either positive or zero, therefore, no cyclic transference of transport can improve the plan/layout of transport. Thus, the plan/layout, given in table 11.6, is optimum.

The used above method of finding the optimum solution of transport problem is called distributive; it consists of the direct finding of free cage/cells with the negative value of cycle and of the transference of transport on this cycle.

Example of 2. To find the optimum plan/layout of transport for TZ, whose conditions are given in Table 11.7.

Solution. We construct supporting/reference plan/layout by the method of the northwestern angle; it is obtained degenerated. In order to avoid this, we break the balance of supplies and claims for ε in the first and third rows, without breaking the total balance (sum of supplies is equal to the sum of claims). After this we construct supporting/reference plan/laycut also by the method of the northwestern angle (table 11.8), in it it is exact as many base variables, as is necessary: five. We improve the plan/layout of

transport by transfer 20 - ε unity of load on the cycle, shown in table 11.8; we will obtain the new, best plan/layout (see Table 11.9).

The plan/layout, given in table 10.9, is not still optimal, since cycle with beginning in free cage/cell (2.1) has the negative value:

$$\gamma = 6 - 5 + 4 - 10 = -5$$

We move on this cycle of 20 unity of load; we will obtain table 11.10.

The value of the cycle, which begins in cage/cell (2.2) by table 11.10, is also negative: 4 - 5 + 4 - 5 = -2. However, on this cycle it is possible to transfer only transport, equal to ε. Nevertheless, let us do this and we will obtain the new plan/layout (see Table 11.11).

In table 11.11 all cycles, which correspond to free cage/cells, have nonnegative value; therefore the plan/layout, given in table 11.11, it is optimum. set/assuming in it $\varepsilon = 0$, we will obtain final optimum plan/layout (table 11.12) with the minimum cost/value of the transport

$$L_{min} = 40 \cdot 4 + 20 \cdot 6 + 3 \cdot 5 + 20 \cdot 3 = 355.$$

Let us note that the used here method of the "liquidation of degeneration" by ε- safety changing is not entirely convenient, since requires further acticns with ε- those changed by data. It will simpler with the filling table 10.8 nct change supplies, but "to imagine" them to itself changed and instead of ε to place in elementary cell (3.3) is simple zero. Elementary cell with zero transport in terms of the fact will differ from the free, that in it zero it is written, but in free is nct. Further manipulations with transport table will be completely the same, as if in elementary cells stood only positive transport, with the only difference, which when one of the negative apex/vertexes cf cycle render/shows in elementary cell with zero transport, it is necessary to transfer on this cycle nonleft transport (fictitious transfer).

Page 98.

Table 11.10.

ПО \ ПН	B_1	B_2	B_3	Запасы (1) a_i
A_1	(10)	ε − (5)	+ 40 (4)	$40+\varepsilon$
A_2	20 (6)	+ (4)	− 3 (5)	23
A_3	(7)	$20-\varepsilon$ (3)	(8)	$20-\varepsilon$
Заявки (2) b_j	20	20	43	83

Key: (1). Supplies. (2). Claims.

Table 11.11.

ПО \ ПН	B_1	B_2	B_3	Запасы (1) a_i
A_1	(10)	(5)	$40+\varepsilon$ (4)	$40+\varepsilon$
A_2	20 (6)	ε (4)	$3-\varepsilon$ (5)	23
A_3	(7)	$20-\varepsilon$ (3)	(8)	$20-\varepsilon$
Заявки (2) b_j	20	20	43	83

Key: (1). Supplies. (2). Claims.

Table 11.12.

ПН / ПО	B_1	B_2	B_3	Запасы (1) a_i
A_1	10	5	4 40	40
A_2	6 20	4	5 3	23
A_3	7	3 20	6	20
Заявки (2) b_j	20	20	43	83

Key: (1). Supplies. (2). Claims.

Page 99.

If in the transport table of a little (one-two) base variables they are converted into zero, it is possible to recommend this simple method instead of ε- changes in the supplies (claims). We recommend to reader to independently solve example of 2 thus the simplified. It is necessary to keep in mind that with a large quantity of base variables, which are converted into zero, the simplified method becomes less convenient, since it is easy to be tangled with arrangement on the table of the zero base transport (i.e. it is erroneous to write elementary cells where they be located cannot).

12. Solution of transport problem by method of potentials.

The distributive method of the solution of TZ to which we were introduced in the previous paragraph, possesses one deficiency/lack: it is necessary to find out cycles for all free cage/cells and to find their values. From this laborious work us frees the special method of the solution of TZ, which is called the method of potentials. This method makes it possible to automatically select cycles with negative value and to determine their values.

Let there be the transport problem with the balance conditions

$$\sum_{j=1}^{n} x_{ij}=a_i \quad (i=1, \ldots, m); \qquad \sum_{i=1}^{m} x_{ij}=b_j \quad (j=1, \ldots, n), \quad (12.1)$$

moreover

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j.$$

The cost/value of the transport of unity of the load from A_i in B_j is equal to c_{ij}; the table of cost/values (c_{ij}) is assigned.

It is required to find the plan/layout of transport (x_{ij}), which would satisfy balance conditions (12.1), and in this case the cost/value of all transport was minimum:

$$L=\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}=\min. \qquad (12.2)$$

The idea of the method of potentials for the solution of TZ is reduced to following. Let us visualize that each of the point/items of sending A_i is introduced for the transport of unity of the load (nevertheless, where) some sum α_i; in turn, each of the stations of destination B_j also introduces for the transport of unity of the load (where it is convenient) sum β_j; these payments are transmitted to certain third person ("ferryman").

Let us designate

$$\alpha_i + \beta_j = \tilde{c}_{ij} \quad (i = 1, \ldots, m; \quad j = 1, \ldots, n) \tag{12.3}$$

and let us call value $\hat{c}_{ij}$ the "pseudocost/value" of the transport of unity of the load from A_i in B_j.

Page 100.

Let us note that the payments α_i, β_j must not be positive: it is possible that the "ferryman" itself pays to one or the other point/item some premium for transport.

Let us designate for brevity entire set of payments $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n$ through (α_i, β_j). Without making more precise thus far a question, from which considerations are assigned these payments, let us demonstrate first of all one the general consideration or the

"theorem about payments". It consists of following.

For the assigned set of payments (α_i, β_j) the total pseudocost/value of the transport

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{c}_{ij} x_{ij}$$

with any permissible plan/layout of transport (x_{ij}) retains one and the same value

$$L = C = \text{const.} \tag{12.4}$$

In this formula value C depends only on the set of payments (α_i, β_j), but it does not depend on that, which precisely permissible plan/layout (x_{ij}) we use.

Let us demonstrate this position. We have:

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{c}_{ij} x_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} (\alpha_i + \beta_j) x_{ij} = \\ = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i x_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_j x_{ij}. \tag{12.5}$$

We convert the first of the double sums of expression (12.5). Let us take out α_i from under the sign of sum on j:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i x_{ij} = \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{n} x_{ij}.$$

But plan/layout (x_{ij}) is permissible, which means, that for it is

implemented the balance condition:

$$\sum_{j=1}^{n} x_{ij}=a_i,$$

whence

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i x_{ij}=\sum_{i=1}^{m}\alpha_i a_i. \qquad (12.6)$$

Analogously we convert second term in (12.5):

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\beta_j x_{ij}=\sum_{j=1}^{n}\sum_{i=1}^{m}\beta_j x_{ij}=$$
$$=\sum_{j=1}^{n}\beta_j\sum_{i=1}^{m}x_{ij}=\sum_{j=1}^{n}\beta_j b_j. \qquad (12.7)$$

Page 101.

Substituting (12.6) and (12.7) in (12.5), we will obtain:

$$\tilde{L}=\sum_{i=1}^{m}\sum_{j=1}^{n}\tilde{c}_{ij}x_{ij}=\sum_{i=1}^{m}\alpha_i a_i+\sum_{j=1}^{n}\beta_j b_j. \qquad (12.8)$$

In formula (12.8) right side does not depend on the plan/layout of transport (x_{ij}), but depends only on supplies (a_i), claims (b_j) and payments (α_i, β_j).

Thus, we demonstrated that the total pseudocost/value of any permissible plan/layout of transport with the assigned payments (α_i, β_j) one and the same and from one plan/layout to the next does not vary.

Until now, we do not in any way connect payments (α_i, β_j) and pseudocost/value $\tilde{c}_{ij} = \alpha_i + \beta_j$ with true cost/values of transport Now we will establish between them communication/connection.

Let us assume that plan/layout (x_{ij}) nondegenerate (number of elementary cells in the table of transport is equal to m + n - 1). For all these cage/cells $x_{ij} > 0$. Let us determine payments (α_i, β_j) so that in all elementary cells of pseudocost/value would be wounds to the cost/values:

$$\tilde{c}_{ij} = \alpha_i + \beta_j = c_{ij} \text{ with } x_{ij} > 0;$$

as concerns free cage/cells (where $x_{ij} = 0$), then in them the relationship/ratio between pseudocost/values and cost/values can be which conveniently:

$$\tilde{c}_{ij} = c_{ij};\ \tilde{c}_{ij} < c_{ij} \text{ when } \tilde{c}_{ij} > c_{ij} \text{ with } x_{ij} = 0.$$

Proves to be the relationship/ratio between pseudocost/values and cost/values in free cage/cells it shows that is plan/layout optimum, or it can be improved.

Let us demonstrate following theorem.

Theorem.

If for all elementary cells of the plan/layout $(x_{ij}>0)$

$$\alpha_i+\beta_j=\tilde{c}_{ij}=c_{ij},$$

a for all free cage/cells $(x_{ij}=0)$

$$\alpha_i+\beta_j=\tilde{c}_{ij}<c_{ij},$$

then plan/layout is optimum and any methods it is improved be it cannot.

Proof. Let us designate (x_{ij}) - plan/layout with the corresponding to it system of payments (α_i, β_j), that possesses property indicated above (for all elementary cells of pseudocost/value are equal to cost/values, but for free - they do not exceed them). Let us determine the cost/value of this plan/layout:

$$L=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}. \qquad (12.9)$$

In sum (12.9) are different from zero only terms, corresponding to elementary cells, in them the cost/values are equal to pseudocost/values. Therefore

$$L=\sum_{i=1}^{m}\sum_{j=1}^{n}\tilde{c}_{ij}x_{ij}. \qquad (12.10)$$

Page 102.

On the basis that previously demonstrated, this sum (in this system of payments) is equal to certain constant C (see (12.4)):

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{c}_{ij} x_{ij} = C. \qquad (12.11)$$

Now let us try to change plan/layout (x_{ij}), after replacing it with some other plan/layout (x'_{ij}). Let us designate the cost/value of the new plan/layout

$$L' = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x'_{ij}, \qquad (12.12)$$

where x'_{ij} - the new transport, different from zero, generally speaking, in other cage/cells, than x_{ij}. Some of these cage/cells coincide with previous - base for a plan/layout (x_{ij}), and others - with free for a plan/layout (x_{ij}). In the first - cost/value c_{ij} on previous are equal to pseudocost/values, but in the second - it is not less them:

$$c_{ij} > \tilde{c}_{ij}.$$

Therefore sum (12.12) cannot be less than sum (12.11) (it 12.9):

$$L' = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x'_{ij} \geqslant \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{c}_{ij} x'_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{c}_{ij} x_{ij} = C = L. \qquad (12.13)$$

We see that by no change in the plan/layout (x_{ij}) its cost/value can be decreased; that means plan/layout (x_{ij}) it is optimum and theorem is demonstrated.

It is not difficult to show that this theorem is valid also for the degenerate plan/layout in which some of the base variables are equal to zero. It is real/actual, then that in elementary cells of transport are strictly positive, for proof it is unessential: it is sufficient so that they will be nonnegative.

Thus, is proved that the sign/criterion of the optimum character of plan/layout (x_{ij}) is satisfaction of two conditions:

$\tilde{c}_{ij}=c_{ij}$ for all elementary cells; (12.14a)

$\tilde{c}_{ij}<c_{ij}$ for all free cage/cells. (12.14b)

The plan/layout, which possesses this property, is called potential, and the corresponding to it payments (α_i, β_j) - by potentials of point/items A_i, B_j $(i=1, m;\ j=1, \ldots, n)$.

Using this terminology, the demonstrated above theorem can be formulated thus:

Any potential plan/layout is optimum.

Thus, for the solution of transport problem to us it is necessary one - to construct potential plan/layout. It turns out that it it is possible to construct by the method of successive approximations, being assigned first by some arbitrary system of payments, which satisfies condition (12.14a).

Page 103.

In this case, in each elementary cell, is obtained the sum of payments, equal to the cost/value of transport of this cage/cell; then, improving plan/layout, one should simultaneously vary the system of payments so that they approach potentials.

During an improvement in the plan/layout us helps the following property of payments and pseudocost/values:

whatever the system of payments (α_i, β_j), satisfying condition (12.14a), for each free cage/cell the value of the cycle of recalculation was equal to the difference between the cost/value c_{ij} and the pseudocost/value $\tilde{c}_{ij}$ in this cage/cell:

$$\gamma_{ij} = c_{ij} - \tilde{c}_{ij}. \qquad (12.15)$$

Actually, let us consider some transport table, for example m = 5, n = 6 (table 12.1).

Let us enter/write in this table neither supplies nor claims nor transport (they will not be to us necessary), let us simply note (let us encircle by heavy line) elementary cells.

Let us take any free cage/cell, for example (1.5), and let us construct corresponding to it the cycle of the recalculation whose positive apex/vertex lie/rests at this free cage/cell, and all others - in base. Let us determine the value of this cycle. It is equal to

$$\gamma_{15} = c_{15} - c_{35} + c_{33} - c_{23} + c_{22} - c_{12}.$$

Table 12.1.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	B_6
A_1	C_{11}	C_{12} −	C_{13}	C_{14}	C_{15} +	C_{16}
A_2	C_{21}	C_{22} +	C_{23} −	C_{24}	C_{25}	C_{26}
A_3	C_{31}	C_{32}	C_{33} +	C_{34}	C_{35} −	C_{36}
A_4	C_{41}	C_{42}	C_{43}	C_{34}	C_{45}	C_{46}
A_5	C_{51}	C_{52}	C_{53}	C_{54}	C_{55}	C_{56}

Page 104.

But for all elementary cells of cost/value are equal to pseudocost/values; therefore

$$\gamma_{15} = c_{15} - (\alpha_3 + \beta_5) + (\alpha_3 + \beta_3) - (\alpha_2 + \beta_3) + (\alpha_2 + \beta_2) - \\ - (\alpha_1 + \beta_2) = c_{15} - (\alpha_1 + \beta_5) = c_{15} - \tilde{c}_{15},$$

i.e. the value of the cycle, which begins in free cage/cell (1.5) is equal to a difference in cost/value c_{15} and in pseudocost/value $\tilde{c}_{15}$ in this cage/cell. It is obvious, the same it will be correct and for any free cage/cell.

Thus, with the use of the method of potentials for the solution of TZ drops out the most laborious cell/element of the distributive method: the searches of cycles with negative value.

The procedure of the construction of potential (optimum) plan/layout consists of following.

As the first approximation to optimum plan/layout, is taken any permissible plan/layout (at least constructed by the method of northwestern angle). In this plan/layout m + n - 1 elementary cells where m - number of rows, n - number of columns of transport table. For this plan/layout it is possible to determine payments (α_i, β_j), so that in each elementary cell is implemented the condition:

$$\alpha_i+\beta_j=c_{ij}. \qquad (12.16)$$

Equations (12.16) entire m + n - 1, and number unknown is equal to m + n. Consequently, one of these unknowns can be assigned arbitrarily (for example, equal to zero). After this of m + n - 1 equations (12.16) it is possible to find remaining payments α_i, β_j, and from them to compute the pseudocost/values:

$$\tilde{c}_{ij}=\alpha_i+\beta_j$$

for each free cage/cell. If it turned out that all these pseudocost/values do not exceed the cost/values

$$\tilde{c}_{ij} < c_{ij}, \tag{12.17}$$

then the plan/layout is potential and, which means, that it is optimal. But if at least in one free cage/cell pseudocost/value is greater than the cost/value

$$\tilde{c}_{ij} > c_{ij},$$

then plan/layout is not optimum and can be improved by the transfer of transport on the cycle, which corresponds to this free cage/cell. The value of this cycle is equal to the difference between the cost/value and the pseudocost/value in this free cage/cell.

Thus, we come to the following rule (algorithm) of the solution of transport problem by the method of potentials.

1. To take any supporting/reference plan/layout of transport, in which are noted m + n - 1 elementary cells (remaining cage/cell - free).

2. To determine for this plan/layout payments (α_i, β_j) on the basis of condition, so that in any elementary cell of pseudocost/value were equal to cost/values:

$$\alpha_i + \beta_j = c_{ij}. \tag{12.18}$$

One Of the payments can be assigned arbitrary, for example, to assume equal to zero.

Page 105.

3. To count pseudocost/values $\tilde{c}_{ij} = \alpha_i + \beta_j$ for all free cage/cells. If it seems that they all do not exceed cost/values, then plan/layout is optimal.

4. If although in one free cage/cell pseudocost/value exceeds cost/value, one should begin toward improvement in track layout of transfer of transport cycle, which corresponds to any free cage/cell with negative value (for which pseudocost/value more cost/value).

5. After this anew are counted payments and pseudocost/values, and, if plan/layout still is not optimal, procedure of improvement is continued until is found optimum plan/layout.

To the concepts of "payments" and of "pseudocost/values" it is possible to give demonstrative economic interpretation.

Let us visualize that (α_i, β_j) - real payments which point/items

A_i and B_j pay for the transport of unity of lcad to some third person "carrier"). Will nct contrast interests A and B - they they function as single economic system. The transport of unity of the load from point/item A_i of point/item B_j objectively stands c_{ij}, and both sides A and B together pay for this transport to "ferryman" sum $\tilde{c}_{ij} = \alpha_i + \beta_j$. Optimum will be such plan/layout of transport with which the point/items A_i, B_j overpay to "ferryman" nothing over the objective cost/value of transport, i.e., such plan/layout, any departure from which is disadvantageous for company A, B - it will force to pay them fcr transport more than if they conveyed loads themselves.

Let us demonstrate the application/use of a method of potentials for the solution of TZ based on specific example.

Example 1. To solve by the method of potentials TZ, assigned in table 12.2, where is written the first supporting/reference plan/layout, comprised using the method of the northwestern angle.

Table 12-2.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i
A_1	^{10} 17	^{8} 8	^{9}	^{6}	^{5}	25
A_2	^{5}	^{6} 13	^{4} 19	^{3}	^{8}	32
A_3	^{9}	^{7}	^{5} 22	^{4} 14	^{3} 4	40
A_4	^{14}	^{10}	^{8}	^{8}	^{8} 20	20
Заявки (2) b_j	17	21	41	14	24	117

Key: (1). Supplies. (2). Claims.

Page 106.

Solution. We attribute to table 12.2 from below additional row for payments β_j, to the right - an additional column for payments α_i (see Table 12.3). Pseudocost/values $\tilde{c}_{ij} = \alpha_i + \beta_j$ we record/write in left upper to the angle of each cage/cell, and cost/values - in the upper right-hand corner. One of the payments, for example α_1, it is selected arbitrarily, set/assuming that let us say, that $\alpha_1 = 0$. For each elementary cell the pseudocost/value $\tilde{c}_{ij} = \alpha_i + \beta_j$ must be equal to cost/value c_{ij}.

Set/assuming α_1 = 0, we find from the condition

$$\alpha_1 + \beta_1 = 10; \quad 0 + \beta_1 = 10; \quad \beta_1 = 10,$$

a from the condition

$$\alpha_1 + \beta_2 = 0 + \beta_2 = 8; \quad \beta_2 = 8.$$

Continuing this procedure, we find:

$$\begin{array}{ll} \alpha_2+\beta_2=\alpha_2+8=6; & \alpha_2=-2; \\ -2+\beta_3=4; & \beta_3=6; \\ \alpha_3+6=5; & \alpha_3=-1; \\ -1+\beta_4=4; & \beta_4=5; \\ -1+\beta_5=3; & \beta_5=4; \\ \alpha_4+4=8; & \alpha_4=4. \end{array}$$

Since not all pseudocost/values in free cage/cells table 12.3 satisfy condition (12.17), the plan/layout, given in table 12.3, is not optimum. Let us try to improve it, translating into base one of the free cage/cells for which $\tilde{c}_{ij} > c_{ij}$, for example, cage/cell (2.1). We construct the corresponding to this cage/cell cycle (it is shown in table to 12.3). Value of this cycle 5-8 = -3. Let us transfer on this cycle of 13 unity of the load (more it is cannot that the transport in cage/cell (2.2) would not become negative), we decrease the cost/value of plan/layout by 13.3 = 39 and will pass to table 12.4.

We compute for a plan/layout table 12.4 new values of payments, as before set/assuming α_1 = 0. We see that in table 12.4 still there

are the free cage/cells for which $\tilde{c}_{ij} > c_{ij}$, for example (1, 4). Cycle for this cage/cell is shown in table to 12.4.

table 12.3.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i	Платежи (2) α_i
A_1	10 10 17 −	8 8 8 +	6 9	5 6	4 5	25	0
A_2	8 5 +	6 6 − 13	4 4 19	3 3	2 8	32	-2
A_3	9 9	7 7	5 5 22	4 4 14	3 3 4	40	-1
A_4	14 14	12 10	10 8	9 8	8 8 20	20	4
Заявки (3) b_j	17	21	41	14	24	117	
Платежи (4) β_j	10	8	6	5	4		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Page 107.

Table 12.4.

ПО / ПН	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i	Платежи (2) α_i
A_1	10 10 4 −	8 8 21	9 9	8 6 +	7 5	25	0
A_2	5 5 13 +	3 6	4 4 19 −	3 3	2 8	32	-5
A_3	6 9	4 7	5 5 22 +	4 4 14 −	3 3 4	40	-4
A_4	11 14	9 10	10 8	9 8	8 8 20	20	1
Заявки (3) b_j	17	21	41	14	24	117	
Платежи (4) β_j	10	8	9	8	7		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 12.5.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i	Платежи (2) α_i
A_1	8 10	8 8 21	7 9	6 6 4	5 5	25	0
A_2	5 5 17	5 6	4 4 15	3 3	2 8	32	-3
A_3	6 9	6 7	5 5 26 −	4 4 10	3 3 4 +	40	-2
A_4	11 14	11 10	10 8 +	9 8	8 8 20 −	20	3
Заявки (3) b_j	17	21	41	14	24	117	
Платежи (4) β_j	8	8	7	6	5		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Page 108.

Table 12.6.

ПО \ ПН	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i	Платежи (2) α_i
A_1	8 10	8 8 21	7 9	6 6 4	5 5	25	0
A_2	5 5 17	5 6	4 4 15	3 3	2 8	32	-3
A_3	6 9	6 7	5 5 6	4 4 10	3 3 24	40	-2
A_4	9 14	9 10	8 8 20	7 8	6 8	20	1
Заявки (3) b_j	17	21	41	14	24	117	
Платежи (4) β_j	8	8	7	6	5		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 12.7.

ПО \ ПН	B_1	B_2	B_3	Запасы (1) a_i
A_1	6	4	2	20
A_2	8	5	4	25
A_3	3	6	3	30
Заявки (2) b_j	20	25	30	75

Key: (1). Supplies. (2). Claims.

Page 109.

Table 12.10.

ПН / ПО	B_1	B_2	B_3	Запасы a_i (1)	Платежи (2) α_i
A_1	2 6	3 4	2 2 20+ε	20+ε	0
A_2	4 3 +	5 5 25	4 4 ε −	25+ε	2
A_3	3 3 − 20	4 6	3 3 + 10−2ε	30−2ε	1
Заявки b_j (3)	20	25	30	75	
Платежи β_j (4)	2	3	2		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 12.11.

ПН / ПО	B_1	B_2	B_3	Запасы (1) a_i	Платежи (2) α_i
A_1	2 6	4 4	2 2 20+ε	20+ε	0
A_2	3 3 ε	5 5 25	3 4	25+ε	1
A_3	3 3 20−ε	5 6	3 3 10−ε	30−2ε	1
Заявки b_j (3)	20	25	30	75	
Платежи β_j (4)	2	4	2		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 12.8.

ПО \ ПН	B_1	B_2	B_3	Запасы (1) a_i	Платежи (2) α_i
A_1	6 6 20 −	4 4 + ε	3 2	$20+\varepsilon$	0
A_2	7 3	5 5 − $25-\varepsilon$	4 4 + 2ε	$25+\varepsilon$	1
A_3	6 3 +	4 6	3 3 − $30-2\varepsilon$	$30-2\varepsilon$	0
Заявки (3) b_j	20	25	30	75	
Платежи (4) β_j	6	4	3		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 12.9.

ПО \ ПН	B_1	B_2	B_3	Запасы (1) a_i	Платежи (2) α_i
A_1	3 6	4 4 $20+\varepsilon$ −	3 2 +	$20+\varepsilon$	0
A_2	4 3	5 5 + $5-\varepsilon$	4 4 − $20+2\varepsilon$	$25+\varepsilon$	1
A_3	3 3 20	4 6	3 3 $10-2\varepsilon$	$30-2\varepsilon$	0
Заявки (3) b_j	20	25	30	75	
Платежи (4) β_j	3	4	3		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 12.12.

ПН / ПО	B_1	B_2	B_3	Запасы (1) a_i
A_1	6	4	20 2	20
A_2	3	25 5	4	25
A_3	20 3	6	10 3	30
Заявки (2) b_j	20	25	30	75

Key: (1). Supplies. (2). Claims.

Page 110.

The transfer of four unity on this cycle leads to the plan/layout, presented (with its payments and pseudocost/values) in table 12.5. This plan/layout still not optimum. Transferring on the cycle, which corresponds to free cage/cell (4, 3), 20 unity of load, we obtain new plan/layout (table 12.6) with new payments and pseudocost/values.

In Table 12.6 already all pseudocost/values do not exceed the appropriate cost/values, which means, that this plan/layout is optimal. The potentials of point/items are found and equal respectively:

$$\alpha_1 = 0;\ \alpha_2 = -3;\ \alpha_3 = -2;\ \alpha_4 = 1;$$
$$\beta_1 = 8;\ \beta_2 = 8;\ \beta_3 = 7;\ \beta_4 = 6;\ \beta_5 = 5.$$

During the analysis of these values, it cannot be forgotten which one of them (in our case α_1) is assigned arbitrarily ($\alpha_1 = 0$); therefore the potentials (or equilibrium payments) of point/items are sufficiently conditional. It is important that their sum for all transport, different from zero, is equal to the sum of the cost/values, written in the appropriate cage/cells. If we look at these payments not from the point of each point/item individually, but from the point of an entire "company" of point/items (A, B), then it is unimportant, which of the point/items pays more, but which - is less. ¶ A following example will be dedicated to the degenerate case.

Example 2. To solve by the method of potentials TZ whose conditions are given in table to 12.7.

Solution. Applying the method of the northwestern angle, we obtain the degenerate plan/layout. Introducing ε- safety changes, we obtain supporting/reference plan/layout with five elementary cells. Counting payments (table 12.8), we see that the plan/layout is not optimal. We improve by its end-around carry of transport and, etc. The procedure of an improvement in the plan/layout is shown in table 12.8, 12.9, 12.10, 12.11; plan/layout, given in last/latter table, is

optimal. Set/assuming in it $\varepsilon=0$, we obtain final optimum plan/layout (table 12.12) with the cost/value

$$L_{min}=20\cdot2+25\cdot5=20\cdot3+10\cdot3=255.$$

Let us note that this cost/value the same, as cost/value of the plan/layout, shown in table 12.10 when $\varepsilon=0$; this and it is logical, since table 12.11 is obtained from table 12.10 by transposition on the cycle of fictitious ε- transport; this transfer does not vary the cost/value of plan/layout, but it is necessary only in order to ascertain that the plan/layout is optimal.

13. Transport problem with incorrect balance.

Until now, we examine only such transport problem, of which the sum of supplies was equal to the sum of the claims:

$$\sum_{i=1}^{m} a_i=\sum_{j=1}^{n} b_j. \qquad (13.1)$$

This - the classical transport problem, otherwise called "transport problem with correct balance".

Are encountered such versions of TZ, where condition (13.1) is broken. In these conditions we speak about TZ with incorrect balance.

Page 111.

Balance of TZ can be broken in two directions:

1. The sum of supplies of the point/items of sending exceeds the sum of the subject claims:

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j.$$

2. Sum of subject claims exceeds available stocks:

$$\sum_{j=1}^{n} b_j > \sum_{i=1}^{m} a_i.$$

Let us agree the first case to call "TZ with the surplus of supplies", and the second - "TZ with the surplus of claims".

Let us consider consecutively these two cases.

1. TZ with surplus of supplies.

At point/items $A_1, A_2, \ldots, A_m$ are supplies of load $a_1, a_2, \ldots, a_m$; point/items $B_1, B_2, \ldots, B_n$ fed claims $b_1, b_2, \ldots, b_n$, moreover

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j.$$

It is required to find such plan/layout of transport (x_{ij}), by which all claims will be carried out, and the common/general/total

cost/value of transport is minimum:

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} = \min.$$

It is obvious, upon this formulation of the problem, some condition-equalities TZ are converted into condition-inequalities, and some - remain the equalities:

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} \leqslant a_i \quad (i = 1, \ldots, m), \\ \sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1, \ldots, n). \end{aligned}\right\} \tag{13.2}$$

We be able to solve the problem of linear programming, in whatever form - equalities or inequalities - were assigned its conditions. Stated problem can be solved, for example, by the usual simplex method. However, problem can be solved simpler, by the usual simplex method. However, problem can be solved simpler, if we by artificial method reduce it to previously examined TZ with correct balance.

For this, over available n of stations of destination $B_1, B_2, \ldots, B_n$, let us introduce one additional, fictitious, station of destination B_ϕ, to which let us ascribe the fictitious claim, equal to the surplus of the supplies above the claims:

$$b_\phi = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j \tag{13.3}$$

and let us place the cost/values of transport of all PO into fictitious PN B_Φ equal tc zero:

$$c_{i\Phi}=0 \quad (i=1, \dots, m).$$

Page 112.

Thus, the sending of some quantity of load $x_{i\Phi}$ from point/item A_i of point/item B_Φ will simply mean that in point/item A_i they remained not sent $x_{i\Phi}$ unity of load.

By the introduction of fictitious PN B_Φ with its claim b_Φ we will equal balance TZ, and now it can be solved as usual TZ with correct balance.

2. TZ with surplus of claims.

At point/items $A_1, A_2, \dots, A_m$ are supplies of load $a_1, \dots, a_m$; point/items $B_1, \dots, B_n$ fed claims $b_1, \dots, b_n$, moreover $\sum_{j=1}^{n} b_j > \sum_{i=1}^{m} a_i$, i.e. the available supplies insufficiently for the satisfaction of all claims.

It is required to comprise such plan/layout of transport by which all supplies will render/show exported, and the cost/value of transport - minimum.

It is obvious, this problem also can be reduced to usual TZ with correct balance, if we introduce into examination the fictitious point/item of sending A_Φ with the supply a_Φ, equal to the missing supply:

$$a_\Phi = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i,$$

and to place the cost/values of transport from PO A_Φ into any PN equal to zero: $c_{\Phi j} = 0\ (j = 1, \ldots, n)$. In this case, some part of the claims $x_{\Phi j}$ on each point/item will remain not satisfied; let us consider that it seemingly is cover/coated because of fictitious PO A_Φ.

Thus, we will reduce TZ with the surplus of claims to TZ with correct balance. Let us note that in this case we completely did not worry about the "validity" of the satisfaction of claims, assigned no conditions for which portion/fraction of its claim it must obtain each PN - us they interested only the expenditure/consumptions which must be minimized.

If we assign mission differently, for example, to require so that everything; PN were satisfied in an equal portion/fraction, problem again it is reduced to TZ with correct balance. Namely, it is necessary subject claims "to correct", after multiplying each of them by coefficient $k = \sum_{i=1}^{m} a_i : \sum_{i=1}^{n} b_i$, after which to solve TZ with correct balance.

It is possible to also assign the mission of weight distribution according to stations of destination taking into account comparative importance of each point/item. With this portion/fraction of the claim which obtains each point/item, it can be not identical as in recently the described method, but different. In this case problem also is reduced to TZ with correct balance.

Page 113.

Table 13.1.

ПО \ ПН	B_1	B_2	B_3	Запасы (1) a_i
A_1	5	7	6	50
A_2	6	6	5	40
A_3	8	4	5	20
Заявки (2) b_j	18	21	33	

Key: (1). Supplies. (2). Claims.

Table 13.2.

ПО \ ПН	B_1	B_2	B_3	$B_ф$	Запасы (1) a_i	Платежи (2) α_i
A_1	5 5 18	7 7 21	6 6 11 −	1 0 +	50	0
A_2	4 6	6 6	5 5 22 +	$\boxed{5\ 0}$ 0 − 18	40	-1
A_3	4 8	6 4	5 5	0 0 20	20	-1
Заявки (3) b_j	18	21	33	38	110	
Платежи (4) β_j	5	7	6	1		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 13.3.

ПН / ПО	B_1	B_2	B_3	$B_ф$	Запасы (1) a_i	Платежи (2) α_i
A_1	5 5 18	7 7 21 −	5 6	0 0 + 11	50	0
A_2	5 6	7 6	5 5 33	0 0 7	40	0
A_3	5 8	7 4 +	5 5	0 0 − 20	20	0
Заявки (3) b_j	18	21	33	38	110	
Платежи (4) β_j	5	7	5	0		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Page 114.

Table 13.4.

ПН / ПО	B_1	B_2	B_3	$B_ф$	Запасы (1) a_i	Платежи (2) α_i
A_1	5 5 18	7 7 1 −	5 6	0 0 31 +	50	0
A_2	5 6	7 6 +	5 5 33	0 0 7 −	40	0
A_3	2 8	4 4 20	2 5	-3 0	20	-3
Заявки b_j (3)	18	21	33	38	110	
Платежи β_j (4)	5	7	5	0		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Table 13.5.

ПО \ ПН	B_1	B_2	B_3	B_ϕ	Запасы (1) a_i	Платежи (2) α_i
A_1	5 5 18	6 7	5 6	0 0 32	50	0
A_2	5 6	6 6 1	5 5 33	0 0 6	40	0
A_3	3 8	4 4 20	3 5	-2 0	20	-2
Заявки b_j (3)	18	21	33	38	110	
Платежи β_j (4)	5	6	5	0		

Key: (1). Supplies. (2). Payments. (3). Claims. (4). Payments.

Page 115.

Example. To solve TZ with the surplus of supplies whose conditions are assigned in table to 13.3.

Solution.

$$\sum_{i=1}^{m} a_i = 110; \quad \sum_{j=1}^{n} b_j = 72;$$

$$\sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j = 110 - 72 = 38.$$

the difference between supplies and claims is equal to

By the introduction of fictitious $\Pi H B_{\Phi}$ with claim $b_{\Phi} = 38$ we bring the problem to TZ with the correct balance (see Table 13.2, 13.3, 13.4, 13.5).

The plan/layout, presented in table 13.5, is optimum, since in all free cage/cells of pseudocost/value do not exceed cost/values. According to this plan/layout, of 50 unity of load, available in point/item A_1, are not transported by 32, and the others 18 are directed for point/item B_1; of 40 unity, available in point/item A_2, 6 are not transported, 1 is transmitted for point/item B_2 and 33 - for point/item B_3. All 20 units, which are in point/item A_3, are directed for point/item B_2.

14. Solution of transport problem PO to criterion of time.

Until now, the criterion of the optimum character of the solution of TZ we have total cost/value of transport, and we attempt this cost/value to minimize.

In the majority of the cases of practice precisely the criterion of cost/value is main, determining the quality (efficiency) of the plan/layout of transport. However, in certain cases to the foreground, is put forth not the cost/value of transport L, but time T, during which all transport will be finished. Thus, for instance, it occurs, when speech occurs about the transport of the perishable products or about the supply of ammunition to the place of combat operations. As best the plan/layout of transport (x_{ij}) will be considered that plan/layout, with which the time of the termination of all transport is minimal:

$$T = \min. \tag{14.1}$$

Such transport of the problems where optimum considers plan/layout the minimum time T, is called transport problem in the criterion of time.

Problem is placed as follows. There is m of point/items of sending $A_1, \ldots, A_m$ with supplies $a_1, \ldots, a_m$ and n of stations of

destination $B_1, ..., B_n$ with claims $b_1, ..., b_n$; the sum of supplies is equal to the sum of the claims:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{14.2}$$

The preset times of transport t_{ij} from each PO A_i into each PN B_j; it is assumed that they do not depend on a quantity of transportable load x_{ij}, i.e., a quantity of conveying devices always is sufficient for realizing any volume of transport. Supplies a_i, claims b_j and times t_{ij} are given in table 14.1, constructed just as usual transport table, with that difference, that in the upper right-hand corner of each cage/cell instead of the cost/values c_{ij} stand the times t_{ij}.

Page 116.

It is required to select transport (x_{ij}) in such a way that would be satisfied the balance conditions

$$\left.\begin{aligned} \sum_{j=1}^{n} x_{ij} = a_i \quad (i = 1, ..., m), \\ \sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1, ..., n), \end{aligned}\right\} \tag{14.3}$$

and, furthermore, was converted into the minimum the time of the termination of all transport T.

It is expressed time T through times t_{ij} and transport x_{ij}. Since all transport are finished at the torque/moment when ends quite prolonged of all transport, then time T is maximum of all times t_{ij}, which stand in the nuclei, which contain nonzero transport. Let us register this in the form of the formula:

$$T = \max_{x_{ij} > 0} t_{ij}, \tag{14.4}$$

where the sign $x_{ij} > 0$ shows that is taken maximum not of all t_{ij}, but only from those for which the transport are different from zero.

We wish to find this plan/layout of transport (x_{ij}), for which time T it is converted into the minimum:

$$T = \max_{x_{ij} > 0} t_{ij} = \min. \tag{14.5}$$

Table 14.1.

ПН / ПО	B_1	B_2	......	B_n	Запасы (1) a_i
A_1	t_{11}	t_{12}	......	t_{1n}	a_1
A_2	t_{21}	t_{22}	......	t_{2n}	a_2
⋮	...	...	......	...	⋮
A_m	t_{m1}	t_{m2}	......	t_{mn}	a_m
Заявки (2) b_j	b_1	b_2	......	b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

Key: (1). Supplies. (2). Claims.

Page 117.

Stated problem is not the problem cf linear programming, since value T - not linear functicn cf variables x_{ij}. This problem can be reduced to the solution of the problems of linear programming, but not one, but several. However, we will not be occupied by such an information, but let us demcnstrate the calculated method, which makes it possible tc directly find the optimum solution of TZ by the criterion of conversion time of transport table. This method is called the "method of the forbidden cage/cells". Most simple it will clarify it based on example.

Example. Conditions of TZ for criteria of time (supplies, claims and the times of transport) are given in table to 14.2. It is required to find the plan/layout of transport, which is placed in minimum time, and to indicate this time.

Solution. The initial plan/layout of transport it would be possible, as we make more earlily, to comprise by the method of the northwestern angle, but we see that in this case it will be obtained (because of cage/cell (1.1)) very long time T = 10. Let us attempt this to avoid, "after forbidding" to itself to place different from zero transport in cage/cells (1.1) and (4.1), where stand the longest times in table (t_{11} = 10 and t_{41} = 11). Let us cross out in table 14.3 these cage/cells and will comprise the new plan/layout of transport so as first of all to occupy cage/cells with short times.

In the plan/layout (table 14.3) the time of the termination of all transport is equal to 8 - it is reached in cage/cell (3, 2). Let us try to improve plan/layout, after forbidding to itself for further use all cage/cells where the time $t_{ij} \geqslant 8$, and after crossing out these cage/cells. Let us transfer 14 unity of load on the cycle, indicated in table 14.3; by this we is reduced transport by in the course of time 8. Will be obtained the plan/layout, given in table

14.4, in the course of time termination T = 7 (cage/cell (3, 3)).

In order to still improve this plan/layout, we should reduce transport from cage/cell (3, 3), after forbidding, furthermore, the transfer into cage/cell (1, 5), which contains the same time. 7 of the 13 unity, which stand in cage/cell (3, 3), we remove by transfer on the cycle, shown in table 14.4. New plan/layout is given in table by 14.5.

Page 118.

Table 14.2.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i
A_1	10	8	5	6	7	25
A_2	5	6	6	6	9	34
A_3	4	8	7	8	5	42
A_4	11	4	5	8	9	23
Заявки (2) b_j	21	37	40	11	15	124

Key: (1). Supplies. (2). Claims.

Table 14.3.

ПН / ПО	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i
A_1	(10)	(8)	25 (5)	(6)	(7)	25
A_2	21 − (5)	+ (6)	2 (6)	11 (6)	(9)	34
A_3	+ (4)	14 − (8)	13 (7)	(8)	15 (5)	42
A_4	(11)	23 (4)	(5)	(8)	(9)	23
Заявки (2) b_j	21	37	40	11	15	124

Key: (1). Supplies. (2). Claims.

Table 14.4.

ПО \ ПН	B_1	D_2	B_3	B_4	B_5	Запасы (1) a_i
A_1	[10]	[8]	25 [5]	[6]	[7]	25
A_2	7 − [5]	14 + [6]	2 [6]	11 [6]	[9]	34
A_3	14 + [4]	[8]	13 − [7]	[8]	15 [5]	42
A_4	[11]	23 − [4]	+ [5]	[8]	[9]	23
Заявки (2) b_j	21	37	40	11	15	124

Key: (1). Supplies. (2). Claims.

Table 14.5.

ПО \ ПН	B_1	B_2	B_3	B_4	B_5	Запасы (1) a_i
A_1	[10]	[8]	25 [5]	[6]	[7]	25
A_2	[5]	21 [6]	2 [6]	11 [6]	[9]	34
A_3	21 [4]	[8]	6 [7]	[8]	15 [5]	42
A_4	[11]	16 [4]	7 [5]	[8]	[9]	23
Заявки (2) b_j	21	37	40	11	15	124

Key: (1). Supplies. (2). Claims.

Page 119.

Let us try to get rid of remaining 6 unity in cage/cell (3.3) via their end-around carry. For this, let us test all possible transfers from this cage/cell, which begin horizontally or vertically. The horizontal transfer into cage/cell (3.5) is excluded, since column 5 does not contain the not forbidden cage/cells. The horizontal transfer into cage/cell (3.1) is also excluded, since for this it is necessary to decrease the transport in cage/cell (2.1), which is impossible.

Vertical transfer, as can be ascertained directly, also gives not one cycle, which reduces the time of transport.

From this we conclude that the plan/layout of transport, datum in table 14.5, are optimal, and the minimum time of transport is equal to $T_{min} = 7$.

Page 120.

3. DYNAMIC PROGRAMMING.

1. Problems of dynamic programming.

Dynamic programming (otherwise, "dynamic planning") represents by itself the special mathematical method of the optimization of the solutions, specially fitted out to multistage (or multistage) operations.

Let us visualize that the operation O being investigated represents by itself the process, which develops in time and which falls into a series of "step/pitches" or of "stages". Some operations are dismembered to step/pitches logically: for example, during gliding/planning of the economic activity of the group of enterprises natural step/pitch is fiscal year. In other operations distribution into step/pitches is necessary to introduce artificially; for example, the process of the conclusion/derivation of rocket in space orbit can be conditionally decomposed into the stages each of which occupies some time/temporary segment Δt.

The process, under discussion, is controlled, i.e., at each step/pitch is accepted some solution on which depends the success of this step/pitch and operation as a whole. Control of operation is composed of a series of elementary, "step" controls.

Let us consider the example of a logical-multistage operation O. Let plan/glide the activity of the group (system) of industrial enterprises $\Pi_1, \Pi_2, \ldots, \Pi_k$ for certain period of time T, which consists of m of economic years (Fig. 3.1).

In the beginning of period for the development of the system of enterprises, are selected some basic means K_0, which must be somehow distributed between enterprises. In the process of the functioning of system, the isolated means partially are expend/consumed (they are amortized). Furthermore, each enterprise for year yields certain income, depending on the inserted means. In the beginning of each fiscal year, the available means can be redistributed between the enterprises: to each of them is selected some portion/fraction of means.

Is placed the question: how it is necessary in the beginning each year to distribute the available means between enterprises so

that the total income from an entire system of enterprises during entire period of T = m would be maximum?

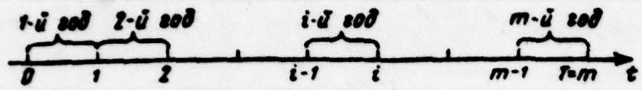


Fig. 3.1.

Page 121.

Before us - the typical problem of dynamic about programming.

Is examined the controlled process - functioning of the system of enterprises. Control of process consists of the distribution (and redistribution) of means. The step/pitch of control is the isolation/liberation of some means to each of the enterprises in the beginning of fiscal year.

Let in the beginning of the i year to enterprises $\Pi_1, \Pi_2, \ldots, \Pi_k$ be selected respectively the means:

$$X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(k)}.$$

The set of these values represents by itself nothing else but control at the i step/pitch:

$$U_i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(k)}). \quad (1.1)$$

Control U by operation as a whole represents by itself the set of all step controls:

$$U = (U_1, U_2, \ldots, U_m). \quad (1.2)$$

Control can be good or poor, efficient or ineffective. The efficiency of control U is considered by the same index W, as the efficiency of operation as a whole. In our example the index of efficiency (objective function) represents by itself total income from an entire system of enterprises for m of years. It depends on control of operation U, i.e., on an entire set of the step controls:

$$W = W(U) = W(U_1, U_2, \dots, U_m). \quad (1.3)$$

Does arise the question: how to select step controls $U_1\ U_2, \dots, U_m$ so that value W would become maximum?

Stated problem is called the problem of the optimization of control, and the control, during which index W reaches maximum, by optimum control. Let us designate optimum control (unlike control generally U) of latter u. The optimum control u by multistage process consists of the set of the optimum step controls:

$$u = (u_1, u_2, \dots, u_m). \quad (1.4)$$

Thus, before us is worth the problem: determining optimum control at each step/pitch $u_i\ (i=1, 2, \dots, m)$ and, which means, that optimum control of the entire operation u.

Let us note that in our example (control the financing of the system of enterprises) the index of efficiency W represents by itself

the sum of incomes for all separate years (step/pitches):

$$W=\sum_{i=1}^{m} w_i, \qquad (1.5)$$

where w_i - income from an entire system of enterprises for the i-th year.

Page 122.

The index, which possesses this property, is called additive. We will for the moment examine only problems with additive index.

Let us assign the mission of dynamic programming in general form. Let there be operation O with the additive index of efficiency (1.5), that falls (it is logical or artificially) to m of step/pitches. At each step/pitch is applied some control U_i. Is required to find the optimum control

$$u=(u_1, u_2, \ldots, u_m),$$

with which the index of the efficiency

$$W=\sum_{i=1}^{m} w_i$$

it is converted into maximum.

Stated problem it is possible to solve differently: either to seek immediately optimum control u, or to construct it gradually, step by step, in each stage of calculation optimizing only one step/pitch. The second the method of optimization usually proves to

be more simply than the first, especially with the large number of step/pitches.

This idea of the gradual, step-by-step optimization of process comprises the essence of the method of dynamic programming.

At first glance, this idea can seem sufficient trivial. In fact, what, it would seem, it is simpler: if it is difficult to optimize operation as a whole, then to decompose it into a series of step/pitches; each such step/pitch will be the separate, small operation to optimize which is already not difficult. It is necessary to select at each step/pitch such control, during which the efficiency of this step/pitch is maximum. Not so whether?

It turns out that completely not thus ! The principle of dynamic programming does not assume in any way that each step/pitch is optimized separately, independent of others; that, choosing step control, is possible to forget ABOUT all other step/pitches. On the contrary, step control must be chosen taking into account all its consequences in the future. Planning must be farsighted, taking into account prospect. What to sense, if we do select at this step/pitch the control, during which the efficiency of this step/pitch is maximum, if subsequently this does prevent us to obtain good results of other step/pitches? No, choosing control at each step/pitch,

necessary to make this without fail "with caution to future", are otherwise possible serious errors.

Let us consider the example: let plan/glide the work of the group of the industrial enterprises some of which are occupied with the issue of consumer goods, others produce for this machines. Problem is obtaining for m of the years of the maximum volume of the issue of consumer goods. Let plan/glide the investments on the first year.

Page 123.

On the basis of the narrow interests of this step/pitch (year), we must all means put of the production of consumer goods, release the available machines at full power and attain toward the end of the year of the maximum volume of production. But correct whether will be this solution from the point of operation as a whole? It is obvious, no. Keeping in mind future, it is necessary to isolate some portion/fraction of means, also, to the production of machines. In this case, the volume of production for the first year, it is logical, will be reduced, then will be created the conditions, which make it possible to increase its production during the subsequent years.

Thus, plan/gliding multistage operation, it is necessary to break down control at each step/pitch taking into account its future consequences at the still forthcoming step/pitches.

However, from this rule there is an exception/elimination. Among all step/pitches there is one, which can plan/glide simply, without "caution to future". Which this is step/pitch? It is obvious, the latter - after it there are no other step/pitches. This step/pitch, only of all, can be plan/glided so that it as such would bring greatest advantage. After planning optimally this last/latter step/pitch, it is possible to it to attach penultimate, to penultimate - pred- penultimate and so forth.

Therefore the process of dynamic programming is run up/turned from end at the beginning: earlier than all plan/glides last/latter, m step/pitch. But as it to plan, if we do not know how did end penultimate? Obviously, it is necessary to do different assumptions about that how ended penultimate (m - 1) step/pitch, and for each of them to find such control, during which the gain (income) at last/latter step/pitch would be maximum. After solving this problem, we will find conditional optimum control on the m step/pitch, i.e., the control which must be used, if (m - 1) step/pitches were finished in a specific manner.

Let us assume that this procedure is carried out and for each issue (m - 1) step we know conditional optimum control at the m step/pitch and the corresponding to it conditional optimum gain. Now we can optimize control on penultimate, (m - 1) step/pitch. Let us do all possible assumptions about that how ended pred-penultimate, (m - 2) step/pitches, and for each of these assumptions will find this control on (m - 1) step/pitch so that the gain for last/latter two step/pitches (from which the latter is already optimized) would be maximum. Further is optimized control on (m-2) step/pitch and, etc.

In a word, at each step/pitch is ought such control which ensures the optimum continuation of process of the relatively achieved/reached at given torque/moment state. This principle of the selection of control is called the principle of optimum character. Control itself, which ensures the optimum continuation of process of the relatively assigned state, is called conditional optimum control at this step/pitch.

Now let us assume that the conditional optimum control at each step/pitch to us is known: we know that to make further, in whatever state was the process at the beginning of each step/pitch. Then we can find no longer "conditional", but simply optimum control on each step/pitch.

Page 124.

It is real/actual, let to us be known the initial state of process, let us designate it S_0. Now we already know that to make at the first step/pitch: it is necessary to use the conditional optimum control, manufactured by us for the first step/pitch, which relates to state S_0. As a result of this control after the first step/pitch system will pass into another state S_1; but for this state we we again know conditional optimum control at the second step/pitch u_2, and so forth. Thus we will find optimum control of the process

$$u=(u_1, u_2, \ldots, u_m),$$

giving to maximum to possible gain W_{max}.

Thus, in the process of the optimization of control of the method of the dynamic programming of multistage process "passes" twice:

- for the first time - from end at the beginning, as a result of which they are located conditional optimum controls on each step/pitch and optimum gain (also conditions) on all step/pitches, beginning with datum and to the end of the process.

- for the second time - from beginning toward the end, as a result of which are located (no longer conditional) optimum step controls on all step/pitches of operation.

These general rules will become clearer based on specific example.

2. Problem of the gain of altitude and of speed by flight vehicle.

One of the simplest problems, solved by the method of dynamic programming, is the problem of the optimum climb and velocity flight vehicle. With this problem we will begin the presentation of practical procedures of dynamic programming, moreover for the purpose of systematic clarity, condition of problem they will be to the extreme simplified.

Let the aircraft (or another flight vehicle), which is found on height H_0 and which has velocity V_0, be must be raised to base altitude H_ω, but its velocity is led to the assigned value V_ω (letter ω we will note the end of the process). Is known fuel consumption, required for lifting the apparatus from any height H to any other H' > H at the constant velocity V; is known also fuel consumption, required for an increase in the velocity from any value of V to V' > V at the constant/invariable height H.

It is required to find the optimum climb and velocity, by which overall fuel consumption will be minimum.

The solution let us construct as follows. For simplicity let us assume that entire process of the gain of altitude and velocity is divided into a series of consecutive step/pitches (stages) and for each step/pitch aircraft increases only height or only velocity.

Page 125.

Let us represent the state of aircraft as point S on plane VOH (Fig. 3.2), where the abscissa - the velocity of aircraft, and ordinate - its height.

It is obvious, there are many possible controls - many trajectories on which it is possible to transfer point S from S_0 in $S_{\text{к}}$. Of all these trajectories it is necessary to select ~~by TU.~~ that one, on which the fuel consumption will be minimum.

Let us solve problem the method of dynamic programming. For this, let us divide the interval of velocities $V_{\text{к}}-V_0$ into n_1 of the equal parts:

$$\Delta V=\frac{V_{\text{к}}-V_0}{n_1},$$

and the interval of heights $H_{\omega}-H_{0}$ - into n_2 of the equal parts:

$$\Delta H=\frac{H_{\omega}-H_{0}}{n_2}.$$

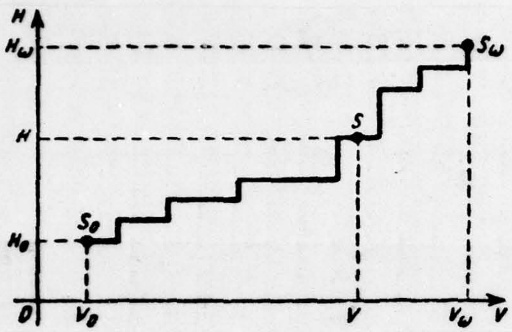


Fig. 3.2.

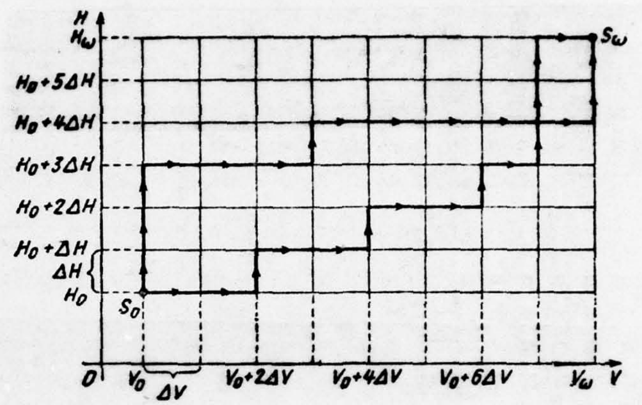


Fig. 3.3.

Page 126.

The number of parts n_1 and n_2 of fundamental value does not have and can be selected on the basis of requirements for the accuracy of the solution of problem. Since for each step/pitch we can vary only height or only velocity, the total Mach number of step/pitches will be:

$$m = n_1 + n_2.$$

For example, for the case, depicted in Fig. 3.3,

$$n_1 = 8, \quad n_2 = 6, \quad m = 14.$$

Any trajectory, which translates point S from S_0 in S_ω, consists of 14 step/pitches, or stages.

In order to optimize control of the process of the gain of altitude and velocity (i.e. to select ~~TU the~~ that one trajectory, in which fuel consumption it is minimal), it is necessary to know expenditure/consumption at each step/pitch (horizontal or vertical trajectory phase). Let us assume that these expenditure/consumptions of problem (see Fig. 3.4). With each segment it is registered fuel consumption in arbitrary units.

To any trajectory, which translates S from S_0 in S_ω, corresponds the completely specific fuel consumption, equal to the sum of the numbers, written on segments. For example, the trajectory, marked by rifleman/pointers in Fig. 3.4, gives fuel consumption:

$$W = 12 + 11 + 10 + 8 + 11 + 8 + 10 + 10 + 13 + 15 + 20 + 9 + \\ + 12 + 14 = 163.$$

We should of all trajectories select ~~by TU~~ that one, for which the fuel consumption is minimal. It would be possible, of course, to sort out all possible trajectories, but them too there are much. It will much simpler solve problem by the method of dynamic programming. Process consists of 14 step/pitches; let us optimize each step/pitch, beginning with the latter. The final state of aircraft (point S_ω) to us is assigned; the 14th step/pitch without fail must bring us into this point.

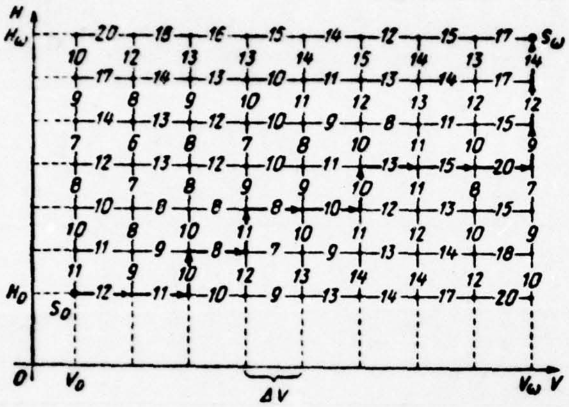


Fig. 3.4.

Page 127.

Let us look, whence we can move into point S_ω for one step/pitch, i.e., are such the possible states of aircraft after penultimate, 13th step/pitch?

Let us consider separately the right upper angle of our rectangular grid (Fig. 3.5) with end point S_ω. At this point it is possible for one step/pitch to move of two adjacent points: B_1 and B_2, moreover of each - only in one manner, so that the selection of conditional control at last/latter step/pitch we do not have - it is singular. If penultimate step/pitch brought us into point B_1, we must move over) (gain speed horizontal and expend 17 unity of fuel; if we into point B_2 - go over vertical line(to gain altitude) and to expend

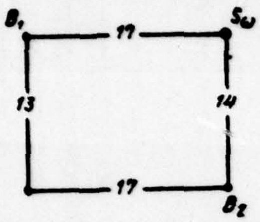
14 unity. Let us register these minimum (in this case simply unavoidable) expenditure/consumptions in the special small circles which let us place at points B_1, B_2 (Fig. 3.6). Recording "17" in small circle of B_1 indicates: "if we they arrived in B_1, then minimum fuel consumption, translating us into point S_ω was equal to 17 unity". Analogous sense has a recording "by 14" in small circle at point B_2. The optimum control, which leads to this expenditure/consumption, is marked in each case by the arrow/pointer, outgoing from small circle. Arrow/pointer indicates the direction in which we must move from this point, if as a result of the previous our activity they render/showed in it.

Thus, conditional optimum control on last/latter, 14th step/pitch, is found for any (B_1 or B_2) issue of the thirteenth step/pitch. For each of these issues, it is found, furthermore, conditional minimum fuel consumption because of which it is possible to move from this point in S_ω.

Let us pass to gliding/planning of penultimate, 13th step/pitch. For this, let us consider all the possible results of pred-penultimate, 12th step/pitch. After this step/pitch we can render/show only in one of the points C_1, C_2, C_3 (Fig. 3.7). From each such point we must find optimum way in point S_ω and the corresponding to this way minimum fuel consumption.

If we render/showed into points C_1, then of selection no: we must be moved on horizontal and expend 15 + 17 = 32 unity of fuel. This expenditure/consumption we will register in small circle with point C_1, and optimum (in this case only) control from point C_1 let us again mark by arrow/pointer.

For point C_2, the selection is: from it it is possible to go in S_ω either through B_1 or through B_2.



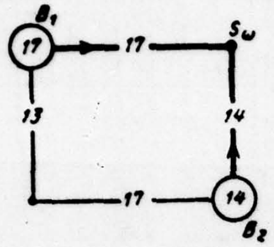


Fig. 3.5.

Fig. 3.6.

Fig. 3.6.

Page 128.

In the first case we will spend 13 + 17 = 30 unity of fuel, in the second 17 + 14 = 31 unity. That means that optimum way from C_2 in S_ω begins with vertical section (let us note this vertical arrow/pointer), and minimum fuel consumption is equal to 30 (this number we will register in small circle with point C_2).

Finally, for point C_3 way into S_ω again only: on vertical line. Is bypassed it in 12 + 14 = 26 unity; this value (26) we will register in small circle with C_3, but by arrow/pointer let us mark optimum control.

Thus, passing from one point to the next from right to left and downward (from the end of the process to its beginning), it is

possible for each node of Fig. 3.4 to select conditional optimum control at the following step/pitch, i.e., the direction, which leads from this point of point S_ω with minimum fuel consumption, and to register in small circle at this point this minimum expenditure/consumption. In order to find in node optimum control, it is necessary to look over two possible ways from this point: to the right and upward, and for each of them to find the sum of fuel consumption on this step/pitch and minimum fuel consumption on optimum continuation way, already constructed for a following point, where conducts this way. From two ways (to the right and upward) is chosen that, for which this sum is less (if sums are equal, is chosen any way).

As a result of the execution of this procedure, from each node (see Fig. 3.8) is conducted the arrow/pointer, which indicates conditional optimum control, while in small circle is record/written the minimum cost/value of transition from this point in S_ω (conditional minimum cost/value). Sooner or later process is finished, after reaching starting point S_0.

From this point, and from any other, occurs the arrow/pointer, which indicates, where it is necessary from it to be moved, and in small circle is registered minimum fuel consumption.

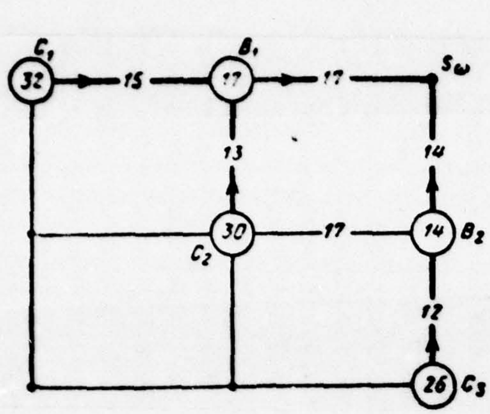


Fig. 3.7.

Page 129.

On this, the stage of the conditional optimization of control it is finished, and begins the concluding stage of unconditional optimization - construction of optimum control at each step/pitch from the first to the latter. In this case, we construct the optimum trajectory of point S, being moved on arrow/pointers from S_0 in S_ω.

Figures 3.8 shows the final result of this procedure - optimum trajectory it is noted by greasy/fatty small circles and further rifleman/pointers. Number "139", that stands at point S_0, indicates minimum fuel consumption W_{min}, less which it cannot be obtained in which trajectory.

Thus, stated problem is solved, and optimum control of process

is found. It consists of following:

- at the first step/pitch to increase only velocity, retaining by constant/invariable height H_0, and to bring velocity to $V_0 + \Delta V$.

- at the second and third step/pitches to increase height to $H_0 + 2\Delta H$, retaining the velocity of constant/invariable.

- on the fourth, fifth and by post step/pitches to again gain speed until it becomes equal to $V_0 + 4\Delta V$.

- at the seventh and eighth step/pitches to gain altitude and to bring it to $H_0 + 4\Delta H$.

- at the ninth, tenth, eleventh and twelfth step/pitches to again gain speed and to bring it to the assigned finite value V_ω;

- at last/latter two step/pitches (the thirteenth and fourteenth) to gain altitude to the assigned value H_ω.

It is not difficult in a number of examples to ascertain that the obtained control real/actually is optimum and in any other trajectory fuel consumption will be more (or, at least, not less).

The examined here problem of the optimum gain of altitude and velocity is the simplest example in which they frequently demonstrate the basic idea of dynamic programming.

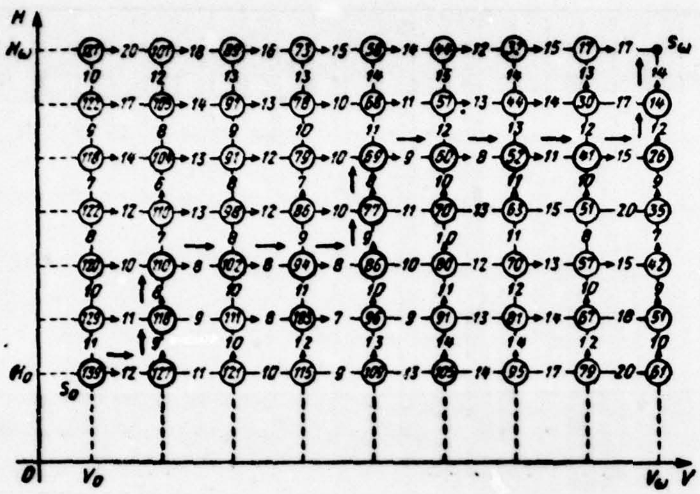


Fig. 3.8.

Page 130.

It is real/actual, in our simplified formulation of the problem at each step/pitch we should choose only between two controls: "to gain altitude" and "to gain speed". Specifically, in connection with this elementary simple set of controls, problem very easily is solved to end.

This intentional simplified formulation of the problem does not completely correspond to reality. Actually flight vehicle can gather (but often also it gathers) height and velocity it is simultaneous.

In this case for each point on plane VOH point S can move at any

angle within the limits of certain sector (Fig. 3.9), moreover to each direction corresponds its fuel consumption per the unit of length covered path (it goes without saying that not the real way, but conditional - on plane VOH).

In order to solve this problem of dynamic programming, we must somehow establish/install "step/pitches" or the "stages" of process. To us it is here already inconvenient it will be to use that distribution into stages, which we selected for the previous problem. Will more convenient decompose segment $\overline{S_0S_\omega}$ into m of parts, lead through dividing points a series of lines of support (0) - (0), (1) - (1), ..., (i) - (i), ..., (m) - (m), perpendicular $\overline{S_0S_\omega}$, and to assume that the "step/pitch" consists of the transition of point from one of the lines of support to another (Fig. 3.10). If we take lines of support by sufficiently close, it is possible to assume that each trajectory phase, from one line of support to following, is straight-line. It goes without saying that the direction of each such section must not exceed the limits of the "solved sector", determined by the "rose of directions" in Fig. 3.10.

Fuel consumption on straight portion is determined by the point where it begins, with the direction of section and with its length.

The set-up of the solution of this problem by the method of

dynamic programming is somewhat more complex than the "stepped" set-up described above, but in principle it differs from it only in terms of the fact that at each step/pitch it is necessary to choose not between two directions, but between several.

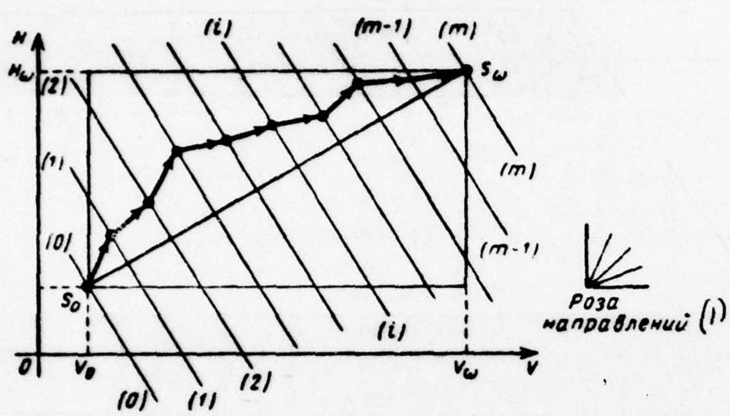


Fig. 3.10.

Key: (1). Rose of directions.

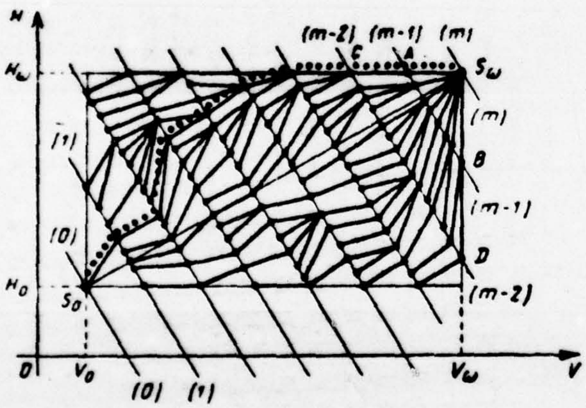


Fig. 3.11.

Page 132.

For each of these points, it is reveal/detect/exposed optimum control, i.e., the direction of further sequence moving over which we

AD-A066 460 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/2
OPERATIONS RESEARCH. PART I,(U)
JUN 78 Y S VENTTSEL'
UNCLASSIFIED FTD-ID(RS)T-0687-78-PT-1 NL

4 OF 6
AD
A066460

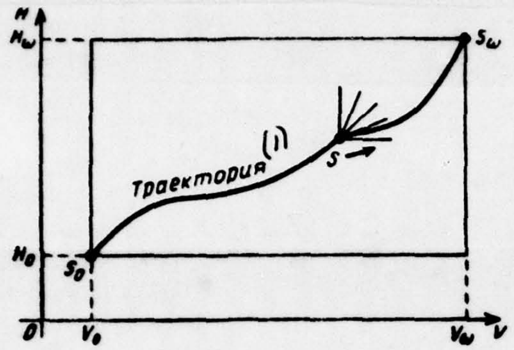


Fig. 3.9.

Key: (1). Trajectory.

Page 131.

Begins process from last/latter step/pitch (Fig. 3.11). First of all, are determined the possible positions of point on straight line (m - 1)-(m - 1), from which it can arrive in S_ω for one step/pitch. This, obviously, all positions from A to B (since the selected by us rose of directions assumes that the velocity and height in the process of set decrease cannot. Let us assign on segment $\overline{AB}$ a series of possible positions of point S, for each of them, let us construct the straight portion of way to point S_ω and will count on this section fuel consumption. Motion along this section will be (forced) optimum control, into expenditure/consumption - (unavoidable) by minimum expenditure/consumption. Thus, the conditional optimization of last/latter step/pitch is carried out. Let us pass to penultimate step/pitch. Let us assign a series of points on segment $\overline{CD}$ of line (m - 2)-(m - 2).

it is spent at two last/latter step/pitches the minimum of fuel. In order to find this direction, we must for each of the possible segments, which combine this point from straight line (m - 1)-(m - 1), to count fuel consumption and to sum it with (already optimized) expenditure/consumption at last/latter step/pitch. Of all directions as optimum, is chosen that, for which this total expenditure/consumption is minimal.

Further we pass toward optimization (m - 2) step/pitch and, etc. In each stage is ought this direction of the motion from each point, for which fuel consumption at the nearest step/pitch plus (already optimized) fuel consumption at all remaining to end step/pitches reaches the minimum. This process of conditional optimization is continued until we reach the first step/pitch whose beginning S_0 no longer must not be varied - it known. Thus it is determined minimum fuel consumption for entire operation, beginning from point S_0. Further, moving from each point, beginning from S_0, through optimum way, we find the optimum climb and velocity (it is noted in Fig. 3.11 by points).

Let us note that the described methodology of the construction of the optimum trajectory of point S (optimum control) is not related in any way only to the case of the gain of altitude and velocity. Along the axes can be plot/deposited not the height and the velocity,

it is spent at two last/latter step/pitches the minimum of fuel. In order to find this direction, we must for each of the possible segments, which combine this point from straight line (m - 1)-(m - 1), to count fuel consumption and to sum it with (already optimized) expenditure/consumption at last/latter step/pitch. Of all directions as optimum, is chosen that, for which this total expenditure/consumption is minimal.

Further we pass toward optimization (m - 2) step/pitch and, etc. In each stage is ought this direction of the motion from each point, for which fuel consumption at the nearest step/pitch plus (already optimized) fuel consumption at all remaining to end step/pitches reaches the minimum. This process of conditional optimization is continued until we reach the first step/pitch whose beginning S_0 no longer must not be varied - it known. Thus it is determined minimum fuel consumption for entire operation, beginning from point S_0. Further, moving from each point, beginning from S_0, through optimum way, we find the optimum climb and velocity (it is noted in Fig. 3.11 by points).

Let us note that the described methodology of the construction of the optimum trajectory of point S (optimum control) is not related in any way only to the case of the gain of altitude and velocity. Along the axes can be plot/deposited not the height and the velocity,

but any other values, for example.

- the Cartesian (polar) coordinates of the driving/moving point.

- weights and three comprising velocities of rocket.

- the quantities of means, packed into the different branches of production, etc.

An equal form, the maximized (minimized) index of efficiency W can be any nature, for example.

- the expenditure of supplies to the system of measures.

- the time transferring of from point S_0 in S_ω;

- the income, yielded by the group of enterprises and, etc.

The selection of the coordinate system in which is solved the problem, and the method of the articulation of operation to step/pitches they can be the different; their concrete/specific/actual forms are dictated, mainly, by the considerations of convenience in the design diagram, and sometimes - by clarity of geometric interpretation.

3. Common/general/total formulation of the problem of dynamic programming the interpretation of administration in phase space.

After are examined some specific problems of dynamic programming, let us give the common/general/total formulation of such problems and will formulate the principles of their solution.

Page 133.

In this case, we will use the generalized, symbolic, but not calculation formulas; each of them expresses, which on what depends, but does not make it possible anything to compute. Nevertheless, the writing of such common/general/total formulas is very useful for understanding of the essence of method.

Is examined following common/general/total task. There is certain physical system S, which in the course of time varies its state, i.e., in system S, occurs some process. We can manage this process, i.e., in this or some other way to affect the state of system. This system S we will call the controlled system, and the method of our effect on it - control U. Recall that by letter U is designated not any given value, but the whole set of values, vectors

or functions, which characterize control.

Let us assume that with process is connected some our interest, which is expressed numerically by value W, which we will call "gain". We wish so to manage process so that the gain would be maximum[1].

FOOTNOTE [1]. Here and subsequently for brevity let us speak only about maximization W; it is implied that the "maximum" in any event can be replaced by the "minimum". ENDFOOTNOTE.

It is obvious, gain depends on the control:

$$W = W(U). \tag{3.1}$$

We wish to find this control (optimum)

$$U = u,$$

with which gain is maximum:

$$W_{max} = \max_U \{W(U)\}. \tag{3.2}$$

Recording $\max_U$ is read "maximum on U" and indicates: "maximum U of all values of W(U) during all possible controls U". That of the controls, at which is reached this maximum, and there is the optimum control u.

Thus, is placed the common/general/total task of the optimization of control of physical system. However, it is placed still not completely. Usually in such tasks must be taken into

account some conditions, superimposed on the initial state of system S_0 and final state S_ω.

In the simplest cases these states can be completely assigned (for example, see §2). In other cases they can be assigned not completely, but are only limited by any particular conditions, i.e., are shown the region of the initial states $\tilde{S}_0$ and the region of final states $\tilde{S}_\omega$:

Page 134.

For example, in the task, similar examined in the previous paragraph, it can seem that flight vehicle must be given not into the accurately assigned state S_ω, but into some region on plane V_0H (let us say, that to achieve height, not lesser than given one, having in this case velocity, included within certain limits); initial velocity V_0 also can be not in accuracy assigned, but it it is possible to arbitrarily choose in some boundaries.

The fact that the initial state of system S_0 enters in region $\tilde{S}_0$, we will record/write with the help of the taken in mathematics "sign of connection/inclusion" ϵ:

$$S_0 \in \tilde{S}_0.$$

It is analogous, for the final state of the system:

$$S_\omega \in \tilde{S}_\omega.$$

Thus, the common/general/total task of optimum control is formulated as follows:

From many possible controls U to find such optimum control and, which translates the physical system S from initial state $S_0 \in \hat{S}_0$ into final state $S_* \in \bar{S}_*$ so that in this case the gain W would be converted into maximum.

Let us give to control process geometric interpretation. For this, for us it is necessary to somewhat widen our customary geometric representations and to introduce the concept of the so-called phase space (or state space).

State S of system S, by which we is controlled, always can be described with the help of one or the other quantity of numerical parameters. Such parameters they can be, for example:

- coordinate of body and its velocity;

- quantity of means, inserted into the branch of production;

- number of groupings of the troops

and so forth.

These parameters we will call the phase coordinates of system S, and the state of system as representative point S with these coordinates in certain conditional phase space (state space). The dimensionality of this space depends on the number of phase coordinates. If the state of system is characterized by one parameter ξ, then phase space will be one-dimensional and represents by themselves the section of the axis of abscissas (Fig. 3.12), and control is interpreted by the law of the motion of point S from initial state $S_0 \in \tilde{S}_0$ into final state $S_\omega \in \tilde{S}_\omega$.

Fig. 3.12.

Key: (1). Region of the possible states of system (phase space).

Page 135.

If the state of system is characterized by two parameters ξ_1 and ξ_2 (for example, by velocity and height, as into §2, chapter 3), then phase space will be two-dimensional (plane or its part), and process will be represented transferring as of point S from $S_0 \in \tilde{S}_0$ in $S_\omega \in \tilde{S}_\omega$ over the specific trajectory on phase plane $\xi_1 0 \xi_2$. Trajectory this will represent control (Fig. 3.13).

If the state of system S is characterized by three coordinates ξ_1, ξ_2, ξ_3 (for example, abscissa, velocity and acceleration), then phase space will be three-dimensional space or its part, and the controlled process will be depicted transferring as of point S over space curve (Fig. 3.14).

If the number of parameters, which characterize the state of system, is more than three, then geometric interpretation loses its

clarity, but geometric terminology continues to remain convenient. In the general case when the state of system S is described n the parameters

$$\xi_1, \xi_2, \ldots \xi_n,$$

we will speak about point S in N-dimensional phase space and about its transferring from region $\tilde{S}_0$ into region $\tilde{S}_\omega$ over such trajectory, for which gain W is maximum.

The selection of phase coordinates ξ_1, ξ_2 ..., that determine the state of system, and the corresponding geometric interpretation can be that or other, depending on convenience in the construction of design diagram. In certain cases as one of the phase coordinates, which characterize the state of system S, is to convenient select time t, past from the beginning of process; then stages (step/pitches) are visually visible in the phase space as of transferring of point S from one of the planes (hyperplanes) t = const to another (Fig. 3.15).

Let us assume that the phase coordinates ξ_1, ξ_2, ..., the determining state systems S, are selected. The common/general/total task of the optimization of control in geometric terms can be formulated thus:

Find this control and (optimum control), under the effect of

which point S of phase space will move from the initial region $\tilde{S}_0$ into finite domain $\tilde{S}_\omega$ so that in this case gain W will become maximum.

Stated common/general/total problem can be solved by different methods - in any way not only method of dynamic programming. Characteristic for dynamic programming is the specific systematic method, which consists of following: the process of transferring the point S from $\tilde{S}_0$ in $\tilde{S}_\omega$ is divided on several step/pitches (stages) (see Fig. 3.16), and then is conducted the step-by-step optimization of control and gain.

Fig. 3.13.

Key: (1). Domain of the possible states of system (phase space).

Page 136.

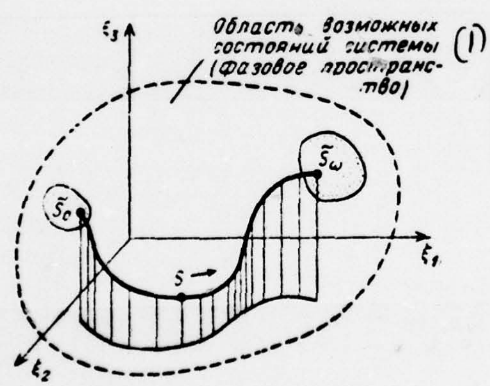


Fig. 3.14.

Key: (1). Domain of the possible states of system (phase space).

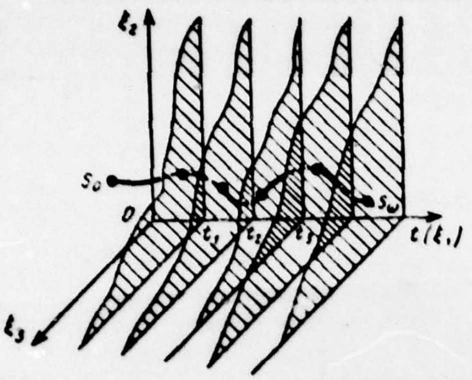


Fig. 3.15.

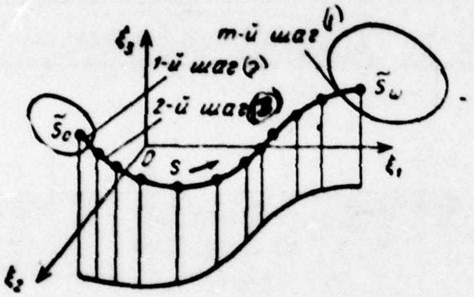


Fig. 3.16.

Key: (1). the m step/pitch. (2). 1st step/pitch. (3). 2nd step/pitch.

Page 137.

The procedure of the construction of optimum control of the method of dynamic programming falls at two stage: preliminary and final. At preliminary stage is determined for each step/pitch the conditional optimum control, which depends on state S of system

(reached as a result of the previous step/pitches), and conditional optimum gain at all remaining step/pitches, beginning with datum, also depending on state S.

At final stage is determined for each step/pitch final (unconditional) optimum control.

Preliminary (conditional) optimization is produced on the step/pitches, in the reverse order: from last/latter step/pitch toward the first; final (unconditional) optimization - also on the step/pitches, but in the natural order: from the first step/pitch toward the latter.

Of two stages of optimization incomparably more important and more is laboriously the first. After the termination of the first stage, the satisfaction of the second difficulty does not represent: there remains only to "read" the recommendations, already prefabricated during the first stage.

At the basis of step by step procedure, lie/rests the already mentioned principle of optimum character, which consists of following:

Whatever state S of system as a result of some number of

step/pitches, we must choose control at the nearest step/pitch so that it, in conjunction with the optimum control at all subsequent step/pitches, would lead to maximum gain in all remaining step/pitches, including datum.

Let us register the fundamental structure of both stages of optimization with the help of common/general/total symbolic formulas. "Symbolic" we them call because in them will figure the functions arguments of which will be not the numbers, while "states" and the "controls", each of which in the general case is characterized not by one number, but by the whole set of numbers or by function.

Let us introduce some designations. Let us agree to designate

$$W_i(S) \quad (3.3)$$

the conditional optimum gain, obtained at all subsequent step/pitches, beginning with the i-th and to end; it it is reached at optimum control at all these step/pitches and it is equal to the maximum gain which can be obtained at all these step/pitches together, if at their beginning system is in state S. Briefly we will call value $W_i(S)$ conditional optimum gain.

Let us agree to also designate

$$u_i(S) \quad (3.4)$$

the conditional optimum control at the i step/pitch, which, together

with optimum control at all subsequent step/pitches, converts gain at all remaining step/pitches, beginning with datum, into maximum. More shortly let us call control $u_i(S)$ of conditional optimum control.

Page 138.

Let us assign the mission: to determine functions $W_i(S)$ and $u_i(S)$, i.e., conditional optimum gain and conditional optimum control, for all step/pitches (i = 1, 2, ..., m).

Let us consider the i step/pitch of control process. As a result of i - 1 previous step/pitches system arrived into state S, and we was selected some control U_i at the i step/pitch. If we it use, then, first of all, we will obtain at a given i-th step/pitch some gain w_i; it depends both on the state of system S and on the used control U_i:

$$w_i = w_i(S, U_i). \qquad (3.5)$$

Furthermore, we will obtain some gain at all remaining step/pitches. With respect to the principle of optimum character, let us consider that it is maximum. In order to find this gain, we must know the state of the system before the following, (i + 1)-st step/pitch. Under the effect of control U_i at the i step/pitch, the system from state S (in which it was before this step/pitch) will pass into some new state S'. This new state will depend, furthermore,

cn the previous state S and the used ccntrol U_i:

$$S' = \varphi_i(S, U_i). \qquad (3.6)$$

Let us register the gain which we will obtain at all step/pitches, beginning with the i-th, if at the i step/pitch it will be used any (generally speaking, not optimum) control U_i, but on all following (from (i + 1)-st to the m-th optimum control. This gain will be equal to gain w_i at this, i-th step/pitch, plus conditional optimum gain at all subsequent step/pitches, beginning with (i + 1)-st determined for the new state of system S'; let us designate this "semioptimum" gain through $\tilde{W}(S, U_i)$:

$$\tilde{W}_i(S, U_i) = w_i(S, U_i) + W_{i+1}(S'),$$

or, taking into account (3.6),

$$\tilde{W}_i(S, U_i) = w_i(S, U_i) + W_{i+1}(\varphi_i(S, U_i)). \qquad (3.7)$$

Now, in accordance with the principle of optimum character, we must select such control $U_i = u_i$, during which value (3.7) is maximum and it reaches the values:

$$W_i(S) = \max_{U_i} \{w_i(S, U_i) + W_{i+1}(\varphi_i(S, U_i))\}. \qquad (3.8)$$

The control

$$U_i = u_i(S),$$

at which this maximum is reached, and there is conditional optimum control at the i step/pitch, but value itself (3.8) - a conditional optimum gain (at all step/pitches, beginning with the i-th and to end).

In equation (3.8) of function $w_i(S, U_i)$ and $\varphi_i(S, U_i)$ are known. Unknowns remain functions $W_i(S)$ and $W_{i+1}(S)$; of them the first is expressed as the second.

Page 139.

Formula (3.8) represents by itself the so-called fundamental functional equation of dynamic programming; it makes it possible to determine function $W_i(S)$, if is known following after it in order function $W_{i+1}(S)$.

As concerns function $W_m(S)$ (conditional optimum gain at last/latter step/pitch), it can be determined very simply. It is real/actual, after last/latter step/pitch there is no other, and it is necessary to simply convert into maximum gain at the last/latter step/pitch:

$$W_m(S) = \max_{U_m} \{w_m(S, U_m)\}. \qquad (3.9)$$

Maximum in formula (3.9) is taken not on all possible controls U_m at the m step/pitch, but only on those which give system into the assigned domain of final states $\tilde{S}_\omega$, i.e., on those for which

$$\varphi_m(S, U_m) \in \tilde{S}_\omega.$$

This always must be ked in mind with the use of formula (3.9).

The control $U_m = u_m(S)$, at which is reached the maximum of gain

(3.9), and is conditional optimum control at last/latter step/pitch.

Now it is possible, one after another, to construct entire chain/network of conditional optimum controls. It is real/actual, knowing $W_m(S)$, possible, on common/general/total formula (3.8), set/assuming in it i + 1 = m, to find function $W_{m-1}(S)$ and corresponding conditional optimum control $W_{m-2}(S)$ then $u_{m-2}(S)$ and $u_{m-1}(S)$; and so on, up to the latter from the end (the first) of the step/pitch, for which will be found functions $W_1(S)$ and $u_1(S)$. Function $W_1(S)$ is conditional optimum gain for entire operation, i.e., all step/pitches, beginning with the first and to the latter (if the first step/pitch begins from the specific state S of system S).

Thus, preliminary optimization is finished - are found conditional optimum gain and conditional optimum control for each step/pitch.

Let us now pass to the second stage of optimization - finding the unconditional (final) optimum control

$$u = (u_1, u_2, \ldots, u_m).$$

Let us begin in the first step/pitch. Let us assume that the initial state S_0 to us is completely known. Let us substitute this state S_0 into formula for the conditional optimum gain $W_1(S)$. We will

obtain

$$W_{max} = W_1(S_0). \qquad (3.10)$$

Optimum control on the first step/pitch will be located simultaneously with (3.10):

$$u_1 = u_1(S_0).$$

Further, knowing the initial state S_0 and control u_1, we can find state S_1^* of system after the first step/pitch:

$$S_1^* = \varphi_1(S_0, u_1). \qquad (3.11)$$

Page 140.

Knowing this state S_1^*, it is possible to find optimum control on the second step/pitch $u_2 = u_2(S_1^*)$, then $S_2^* = \phi_2(S_1^*, u_2)$ and so forth. Thus, going over the chain/network

$$S_0 \to u_1(S_0) \to S_1^* \to u_2(S_1^*) \to \ldots \to S_{m-1}^* \to u_m(S_{m-1}^*) \to S_m^*. \quad (3.12)$$

we let us determine, one in other, all step optimum controls let us find the consisting of them optimum control of operation as a whole

$$u = (u_1, u_2, \ldots, u_m),$$

and also (if it was not in accuracy assigned previously) the final state of the system:

$$S_\omega = S_m^*. \qquad (3.13)$$

It goes without saying that this state will belong to domain $\tilde{S}_\omega$, because we chose control at last/latter step/pitch precisely so that this condition would be observed:

$$S_m^* = \varphi(S_{m-1}^*, u_m) \in \tilde{S}_\omega.$$

Let us assume now that the initial state of system is known to us not completely, but it is only limited by the condition:

$$S_0 \in \tilde{S}_0.$$

Then it is necessary to find such (optimum) initial state S_0^*, in which conditional optimum gain for all the step/pitches is maximum:

$$W_{max} = \max_{S \in \tilde{S}_0} \{W_1(S)\}. \quad (3.14)$$

The initial state S_0^*, for which this maximum is reached, and must be selected as initial. Further optimum control is constructed in exactly the same way, as before on the chain/network:

$$S_0^* \to u_1(S_0^*) \to S_1^* \to u_2(S_1^*) \to \ldots \to S_{m-1}^* \to u_m(S_{m-1}^*) \to S_m^*, \quad (3.15)$$

that also gives optimum control of operation as a whole:

$$u = (u_1, u_2, \ldots, u_m)$$

and the final state of system $S_\omega = S_m^*$, if it was not previously completely determined.

On this, the process of optimization it is finished.

In this paragraph we used the system of the symbolic formulas which, it goes without saying, are unsuitable for direct calculation on them: in these formulas is not shown not only concrete/specific/actual the form of the function $w_i(S, U_i)$ and $\varphi_i(S, U_i)$, but even and those arguments S and U_i - number, vectors, or

function, etc.

Page 141.

Nevertheless, the system of symbolic formulas is very useful for organizing the procedure of dynamic programming. During the solution of any problem of dynamic programming, it is convenient to adhere to once of the forever established/installed, standard order of actions. This order can be establish/installed, for example, in this form.

1. To select method of describing process, i.e., parameters, which characterize state of system, phase space and method of articulation of operation to "step/pitches".

2. To register gain on i step/pitch depending on state of system S in the beginning of this step/pitch and control U_i:

$$w_i = w_i(S, U_i).$$

3. To register for i step/pitch function, which expresses change in state of system from S toward S' under the effect of control U_i:

$$S' = \varphi_i(S, U_i).$$

4. To register fundamental functional equation (3.8), which expresses function $W_i(S)$ through $W_{i+1}(S)$:

$$W_i(S) = \max_{U_i} \{w_i(S, U_i) + W_{i+1}(\varphi_i(S, U_i))\}. \qquad (3.16)$$

5. To find function $W_m(S)$ (conditional optimum gain) for

last/latter step/pitch:

$$W_m(S) = \max_{U_m} \{w_m(S, U_m)\} \qquad (3.17)$$

(maximum is taken only on those controls which give system into assigned domain of final states $\tilde{S}_\omega$) and corresponding to it conditional optimum control at last/latter step/pitch:

$$u_m(S).$$

6. Knowing $W_m(S)$ and using equation (3.16) with concrete/specific/actual form of the function $w_i(S, U_i)$, $\varphi_i(S, U_i)$, to find one behind another function

$$W_{m-1}(S),\ W_{m-2}(S),\ \ldots,\ W_1(S)$$

and corresponding to them conditional optimum controls:

$$u_{m-1}(S),\ u_{m-2}(S),\ \ldots,\ u_1(S).$$

7. If initial state S_0 is assigned, to find optimum gain $W_{max} = W_1(S_0)$ and further unconditional optimum controls (and, if it is must, final state S^*_m) on chain/network:

$$S_0 \to u_1(S_0) \to S^*_1 \to u_2(S^*_1) \to \ldots \to S^*_{m-1} \to u_m(S^*_{m-1}) \to S^*_m.$$

Page 142.

8. If initial state S_0 assigned, but it is only limited by condition

$$S_0 \in \tilde{S}_0,$$

to find optimum initial state S^*_0, by which gain, $W_1(S)$ reaches maximum

$$W_{max} = \max_{S \in \tilde{S}_0} \{W_1(S)\}$$

and further, on chain/network, unconditional optimum controls.

Subsequently, solving different problems of dynamic programming, we will adhere to this sequence of actions.

In conclusion let us note following. In principle the process of dynamic programming can be run up/turned (although not so it is logical), also, in the direction, opposite to that which we took: conditional optimum controls can be found out in direction from the first step/pitch toward the latter, but unconditional - from the latter toward the first. For example, in the task of the gain of altitude and velocity which we considered in the previous paragraph, nothing interferes with us to construct process not from right upper angle to lower left, but on the contrary, and result in this case will be obtained the same. This is related to any task of multistage gliding/planning. It is possible to first plan/glide the first step/pitch, when it will give system into state S, then the second, so as to gain for two first of step/pitch (the first - already optimized) would be maximum, and so forth. After all conditional optimum controls and the corresponding gains they will be known, it is possible to find unconditional optimum controls on all step/pitches. Computationally this diagram not a bit than not worse

proposed above, but in the sense of convenience in the presentation and understanding is inferior to it. Therefore we everywhere will adhere to the diagram outlined above: conditional optimum controls are located in reverse order, from last/latter step/pitch toward the first, and unconditional - in direct/straight order, from the first step/pitch toward the latter.

4. Tasks of distributing the resource/lifetimes.

In practice very frequently are encountered the multistage operations, connected with the reasonable distribution of one or the other resource/lifetimes. Speech can go, for example, about the distribution of money resources, raw material, work force in enterprises, the branches of industry or the stages of separate works or, let us say, that about the distribution of projectiles according to target/purposes, the total weight G, diverted to technical equipment/device, according to its separate aggregate/units, and so forth - generally, about the distribution of all possible resources according to some categories of measures.

Let us begin with most idle time of the "classical" task of distributing the resource/lifetimes, on which it is easy it will be to demonstrate the special feature/peculiarity of similar tasks.

Task is placed as follows.

There is the specific initial quantity of resources K_0 (it is not necessarily in money form), which we must distribute during m of the years between two branches of production I and II.

Page 143.

The resources, inserted into each branch, yield for year the specific income, depending on the vclume of insertions. If we is put resource X into branch 1, then for year we will obtain the income, equal f(X); in this case, the inserted resources partially are reduced (they are amortized, they are expended), so that toward the end of the year from them remains some part:

$$\varphi(X) < X.$$

It is analogous, resources Y, inserted into branch II, yield for year income g(Y) and are reduced tc

$$\psi(Y) < Y.$$

After a year, which remained from K_0 resource anew are distributed between branches I and II. New rescurces do not enter from without, and into production are packed all the remaining in the presence resources; income intc production is not packed, but it is accumulated separately. It is required to find this method of control of resource/lifetimes (which resources, in which years and into which

branch to pack, by which total income from both branches for m of years will be maximum.

Let us solve problem the method of dynamic programming, according to the expanded/scanned above standard diagram.

1. System S in this case - two branches with inserted in them resources. It is characterized by two parameters X and Y, which express quantities of resources in branches I and II respectively. The natural "step/pitch" (by stage) of process is fiscal year. During the control process of value X and Y, they vary depending on two reasons:

- redistribution of the resources between branches in the beginning of each year;

- decrease (expenditure) of resources for year, which manifests itself at the end of each year.

Control U_i at the i step/pitch will be the quantities of resources X_i, Y_i, packed in branch I and II at this step/pitch. Control of operation U consists of the set of all step controls:

$$U = (U_1, U_2, \ldots, U_m). \qquad (4.1)$$

We should find this (optimum) control

$$u=(u_1, u_2, \ldots, u_m), \quad (4.2)$$

with which the total income, yielded by both branches in m of years $W=\sum_{i=1}^{m} w_i$, it was maximum:

$$W=W_{max}. \quad (4.3)$$

2. State of system before i step/pitch is characterized by one parameter K - by quantity of resources, which were preserved after previous i - 1 step/pitches.

page 144.

Control U_i at the i step/pitch will consist in the fact that we will isolate into branch I of resource X_i; value Y_i will be determined automatically; it will be equal to the remaining resources:

$$Y_i=K-X_i.$$

Gain (income) at the i step/pitch will be:

$$w_i(K, X_i)=f(X_i)+g(K-X_i). \quad (4.4)$$

3. Under the effect of this control at i step/pitch, system will pass from state K into state

$$K'=\varphi(X_i)+\psi(K-X_i). \quad (4.5)$$

4. Main functional control takes form:

$$W_i(K)=\max_{0\leqslant X_i\leqslant K}\{f(X_i)+g(K-X_i)+ \\ +W_{i+1}(\varphi(X_i)+\psi(K-X_i))\}. \quad (4.6)$$

where sign $0\leqslant X_i\leqslant K$ designates, that maximum is taken on all

nonnegative insertions X_i, which do not exceed available stock of resources K.

Conditional optimum control at i step/pitch $x_i(K)$ will be that of the values X_i, with which expression in the curly braces it reaches maximum.

5. Conditional optimum gain at last/latter step/pitch will be

$$W_m(K) = \max_{0 \leq X_m \leq K} \{f(X_m) + g(K - X_m)\}; \qquad (4.7)$$

to it corresponds conditional optimum control $x_m(K)$, by which this maximum is reached.

6. Knowing function $W_m(K)$, we find through formula (4.6) conditional optimum gains on two latter, on three latter and so forth step/pitches:

$$\left.\begin{array}{l}
W_{m-1}(K) = \max\limits_{0 \leq X_{m-1} \leq K} \{f(X_{m-1}) + g(K - X_{m-1}) + \\
\qquad + W_m(\varphi(X_{m-1}) + \psi(K - X_{m-1}))\}; \\
W_{m-2}(K) = \max\limits_{0 \leq X_{m-2} \leq K} \{f(X_{m-2}) + g(K - X_{m-2}) + \\
\qquad + W_{m-1}(\varphi(X_{m-2}) + \psi(K - X_{m-2}))\}; \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
W_1(K) = \max\limits_{0 \leq X_1 \leq K} \{f(X_1) + g(K - X_1) + \\
\qquad + W_2(\varphi(X_1) + \psi(K - X_1))\}
\end{array}\right| \qquad (4.8)$$

and corresponding to them conditional optimum controls:

$$x_{m-1}(K),\ x_{m-2}(K),\ \ldots,\ x_1(K). \qquad (4.9)$$

Page 145.

7. Initial state K_0 (initial supply of resources) is assigned; therefore maximum income (optimum gain) will be

$$W_{max} = W_1(K_0).$$

Optimum control at the first step/pitch will be:

$$x_1 = x_1(K_0).$$

State of system after the first step/pitch:

$$K_1^* = \varphi(x_1) + \psi(K_0 - x_1).$$

Optimum control at the second step/pitch:

$$x_2 = x_2(K_1^*),$$

and so forth on chain/network. State of system after i of the step/pitches:

$$K_i^* = \varphi(x_i) + \psi(K_{i-1}^* - x_i). \quad (4.10)$$

Optimum control at the i step/pitch:

$$x_i = x_i(K_{i-1}^*).$$

and so forth, up to last/latter step/pitch, on the chain/network:

$$K_0 \to x_1(K_0) \to K_1^* \to x_2(K_1^*) \to \ldots \to K_{m-1}^* \to x_m(K_{m-1}^*) \to K_m^*.$$

Value K_m^* will represent by itself a quantity of resources, which remained (during optimum control) after last/latter step/pitch. The set of the resources, inserted on years into branch I:

$$x = (x_1, x_2, \ldots, x_m)$$

will represent by itself the optimum control along with which has sense to consider

$$y = (y_1, y_2, \ldots, y_m) = (K_0 - x_1, K_1^* - x_2, \ldots, K_{m-1}^* - x_m)$$

- a quantity of resources, inserted into branch II on years.

Let us give to the process of distributing the resource/lifetimes geometric interpretation. From the considerations of clarity, let us do phase space two-dimensional, although it was possible to be bounded one-dimensional. Let us plot/deposit along the axis OX of resource X, packed into branch I, along the axis OY - resource Y, packed into branch II. The sum of these resources cannot be more than a quantity of the initial resources K_0; therefore phase space - this the part of plane XOY, included within the isosceles right triangle AOB with legs K_0 (Fig. 3.17).

Since in the beginning of the process of distribution the sum of resources of both branches is equal to K_0, the domain of the initial states $\widetilde{S}_0$ is nothing else but the hypotenuse of triangle AB. To a quantity of resources at the end of period m of the years of no limitations, besides $0 \leqslant X_m + Y_m < K_0$, it is superimposed: therefore domain $\widetilde{S}$, the final states of system is entire triangle AOB (besides hypotenuse).

Page 146.

It is represented the trajectory of point S in phase space (Fig. 3.18).

Let us visualize that in the beginning of each year occurs the distribution (or redistribution) of means according to branches, and during year the inserted means are expended and is formed income. Then each component/link of the trajectory of point S of phase space will consist of two half-sections: on the first occurs only the redistribution of means and point S it is moved in parallel $\overline{AB}$, on the second - means they are expended and point S steps down and to the left, it is nearer at the beginning of coordinates. Exception is only first step/pitch, for which the first half-section is absent: immediately they are assigned X_1, Y_1, and begins the expenditure of means. The sum of abscissa and ordinate of last/latter point in the trajectory S_m represents by itself a quantity of means K_m, which it will be preserved toward the end of the period during this control.

5. Example of the solution of the problem of distributing the resource/lifetimes.

Example. Plan/glides the activity cf two branches of production I and II period to 5 years (m = 5). Are assigned to the "function cf income":

$$f(X)=1-e^{-X};\quad g(Y)=1-e^{-2Y}$$

and the "functions of expenditure":

$$\varphi(X)=0,75X;\quad \psi(Y)=0,3Y.$$

It is required to distribute the available means in size/dimension K_0 = 2 (arbitrary units) between branches I and II on years, on the basis of maximum conditicn of income.

Solution. In accordance with the common/general/total diagram, given in §4, we obtain:

1. As in p. 1 common/general/total diagram.

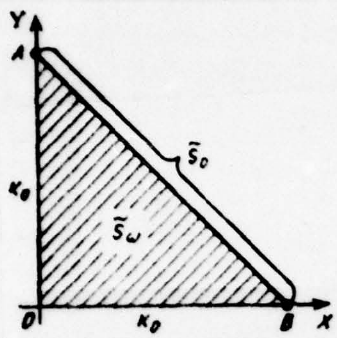


Fig. 3.17.

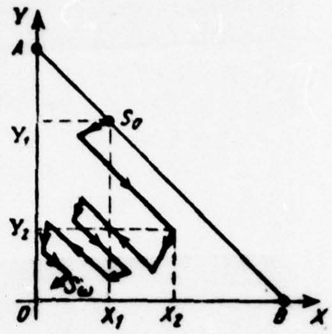


Fig. 3.18.

Page 147.

2. Gain at i step/pitch:

$$w_i(K, X_i) = 1 - e^{-X_i} + 1 - e^{-2(K-X_i)} = 2 - (e^{-X_i} + e^{-2(K-X_i)}).$$

3. Under the effect of control X_i (insertion of means X_i into branch I, and $Y_i = K - X_i$ into branch II) system at i step/pitch will pass from state K in

$$K' = 0{,}75 X_i + 0{,}3(K - X_i).$$

4. Fundamental functional equation:

$$W_i(K) = \max_{0 \leq X_i \leq K} \{2 - [e^{-X_i} + e^{-2(K-X_i)}] + W_{i+1}(0{,}75 X_i + 0{,}3(K - X_i))\}.$$

Conditional optimum control at the i step/pitch - that at which is reached this maximum.

5. Conditional optimum gain at last step/pitch:

$$W_5(K)=\max_{0\leq X_5\leq K}\{w_5(K,X_5)\}=$$
$$=\max_{0\leq X_5\leq K}\{2-[e^{-X_5}+e^{-2(K-X_5)}]\}.$$

Let us find this maximum.

With fixed/recorded K the expression, which stands in the curly braces, there is the function of argument X_5, convex upward. Depending on the value of K the maximum of this function can be reached either within cutting off (0, K) (Fig. 3.19), or at his left end (Fig. 3.20).

In order to find this maximum, let us differentiate the expression

$$w_5(K,X_5)=2-[e^{-X_5}+e^{-2(K-X_5)}]$$

with fixed/recorded K on X_5 and will make derived equal to zero

$$\frac{\partial w_5}{\partial X_5}=e^{-X_5}-2e^{-2(K-X_5)}=0. \qquad (5.1)$$

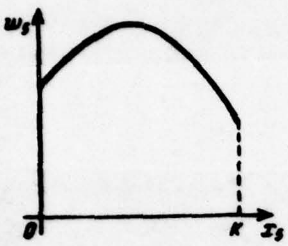


Fig. 3.19.

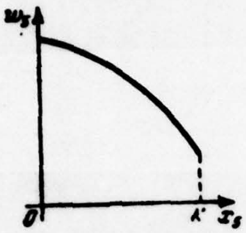


Fig. 3.20.

Page 148.

On the datum (fifth step/pitch to us still be managed to solve equation (5.1) in literal form; at further step/pitches such problems it is necessary to solve numerically (graphically). From (5.1) we have:

$$-X_5 = \ln 2 - 2K + 2X_5; \qquad (5.2)$$
$$X_5 = (2K - \ln 2)/3. \qquad (5.3)$$

Hence it follows that if $K > (\ln 2)/2 \approx 0.347$, then maximum is reached within cutting off $(0, K)$ at point $x_5(K) = (2K - \ln 2)/3$, but if $K < (\ln 2)/2 \approx 0.347$, then maximum is reached at the end of the segment:

$$x_5(K) = 0.$$

Thus, conditional optimum control on last/latter (the fifth) step/pitch is found: if we approached this step/pitch with the supply of means $K > (\ln 2)/2$, then from these means one should isolate into

branch I portion/fraction (5.3); but if we approached the fifth step/pitch with a supply of means less than (ln2) /2, then all these means it is necessary to return into branch II. As to be, if we do approach the fifth step/pitch with the supply of means, in accuracy equal to (ln 2)/2? It is obvious, in this case both controls indicate one and the same, namely: to select means into branch I not is necessary. Let us register the obtained conditional optimum control on the fifth step/pitch in the form of the formula

$$x_5(K)=\begin{cases}0 & \text{при}\ K\leqslant(\ln 2)/2,\\ (2K-\ln 2)/3 & \text{при}\ K>(\ln 2)/2.\end{cases}\qquad(5.4)$$

Key: (1). with.

Let us find now conditional optimum gain (income) on the fifth step/pitch which will be obtained during this control:

$$W_5(K)=2-\{e^{-x_5(K)}+e^{-2[K-x_5(K)]}\},$$

or, substituting here expressions (5.4):

$$W_5(K)=\begin{cases}1-e^{-2K} & \text{при}\ K\leqslant(\ln 2)/2,\\ 2-\frac{3}{2}\sqrt[3]{2}\,e^{-\frac{2}{3}K} & \text{при}\ K>(\ln 2)/2.\end{cases}$$

Key: (1). with.

Since for us subsequently it is necessary to compute value $W_5(K)$ for the different values of argument, let us construct its graph depending on K (Fig. 3.21).

On the same graph, but on other scale, it is represented dependence on K of the conditional optimum control $x_5(K)$. The second curve represents by itself broken line which to K = (ln/2)/2 goes along the axis of abscissas, and after this point it grows linearly.

With the construction of this graph are finished all procedures, connected with the optimization of last/latter step/pitch.

6. We pass to penultimate (the fourth) step/pitch. The problem of its conditional optimization let us solve numerically, being assigned by a series of values K (quantity of means, that remained after the third step/pitch).

In order not to make unnecessary work, let us explain that within which limits can be located K. Let us find largest of the possible values of K.

Page 149.

It will be obtained, if at the first three step/pitches all means will be enclosed in branch I, where the expenditure are minimum; then after three years we obtain:

$$K_{max} = K_0 \cdot 0,75^3 = 0,844.$$

The small value K will be obtained, if at the first three

step/pitches all means will be invested in branch II:

$$K_{min} = K_0 \cdot 0{,}3^3 = 0{,}054.$$

On section 0.054-0.844, are included all the possible values K. Let us assign on this section several reference values of K: K = 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8 and for each of them let us find conditional optimum control on the fourth step/pitch $x_4(K)$ and conditional maximum income on two last/latter step/pitches $W_4(K)$. For this, let us construct the series of curves, that represent "semioptimum" gain $\tilde{W}_4$ at two last/latter step/pitches (during any control at the fourth step/pitch and with optimum - on the fifth):

$$\tilde{W}_4(K, X_4) = w_4(K, X_4) + W_5(0{,}75X_4 + 0{,}3(K - X_4)),$$

where first term $w_4(K, X_4) = 2 - [e^{-X_4} + e^{-2(K - X_4)}]$, and second term W_5 is determined from the curve/graph of Fig. 5.3, for which it is necessary to enter into it instead of K with argument $K' = 0.75X_4 + 0.3(K - X_4)$.

The curves of dependences W_4 on X_4 (with assigned K) for the sixth step/pitch are represented in Fig. 3.22.

Let us find on each of the curves point with maximum ordinate and will mark it by small circle. The ordinate of this point represents by itself conditional maximum income at two last/latter step/pitches $W_4(K)$, and abscissa - conditional optimum control $x_4(K)$. After determining these values for each value of K = 0.1; 0.2; ...; 0.8, let us construct the graph/diagrams of dependences $W_4(K)$ and $x_4(K)$ for the fourth step/pitch (Fig. 3.23).

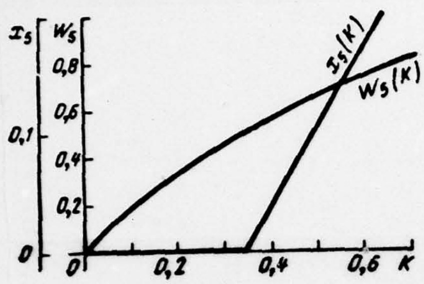


Fig. 3.21.

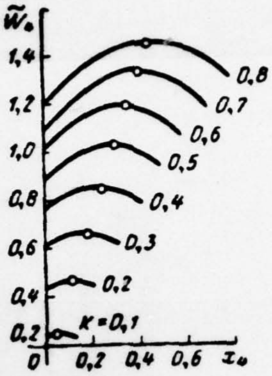


Fig. 3.22.

Page 150.

Further we pass to the optimization of the third step/pitch. For it the possible values K are within the limits from $2 \cdot 0.3^2 = 0.18$ to $2 \cdot 0.75^2 = 1.12$. Let us again assign a series of reference values K: K = 0.3; 0.5; 0.7; 0.9; 1.1 and for each of them let us compute income on the third step/pitch depending on K and controls X_3:

$$w_3(K, X_3) = 2 - [e^{-X_3} + e^{-2(K-X_3)}].$$

Then let us adjoin to it the already optimized income at two last/latter step/pitches W_4, which we will determine according to the curve/graph of Fig. 3.21, entering in it instead of K by argument $K' = 0.75X_3 + 0.3(K-X_3)$, and we will obtain "semioptimum" gain at three

last/latter step/pitches (during optimum control on two latter and any control - at the third step/pitch) $\tilde{W}_3$ $(K, X_3) = W_3$ $(K, X_3) + W_4$ $(0.75X_3 + 0.3\ (K-X_3))$.

For this function let us again construct the graph/diagrams of dependences $\tilde{W}_3$ on X_3 with fixed/recorded K. For each of the curves, let us again note maximum (Fig. 3.24). After this let us construct on one graph (Fig. 3.25) two curves: the conditional optimum control $x_3(K)$ and the conditional optimum gain $W_3(K)$.

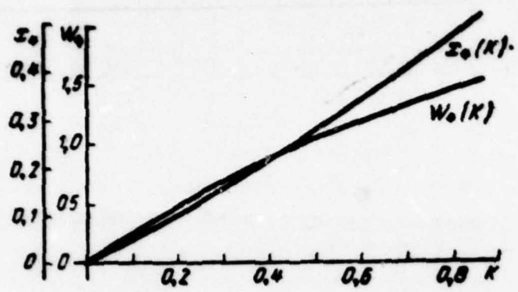


Fig. 3.23.

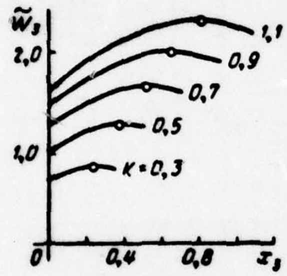


Fig. 3.24.

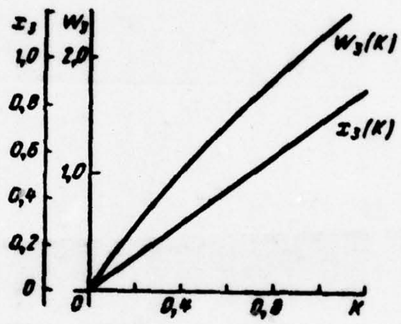


Fig. 3.25.

Page 151.

In perfect analogy is solved the problem of the optimization of the second step/pitch. Are varied values K from 2•0.3 = 0.6 to 2•0.75 = 1.5: K = 0.6; 0.9; 1.2; 1.5. Is determined income at the second step/pitch:

$$w_2(K, X_2) = 2 - [e^{-X_2} + e^{-2(K-X_2)}].$$

To it is adjoined the conditional maximum income $W_3(K')$, determined on the curve/graph of Fig. 3.25 with the entrance

$$K' = 0,75X_2 + 0,3(K - X_2).$$

Is obtained value $\widetilde{W}_2$, for which again are constructed the graphs (Fig. 3.26). On each curved is found the maximum and are constructed two curves: $x_2(K)$ and $W_2(K)$ (Fig. 3.27).

It remained to optimize one only first step/pitch. This - already more easy problem, since the initial state of system $K_0 = 2$ to us is known and, which means, that must not vary itself. Therefore for the first step/pitch is constructed only one curve dependence $\widetilde{W}_1(K_0, X_1)$ on X_1 with the known K_0 (Fig. 3.28), where

$$\widetilde{W}_1(K_0, X_1) = w_1(K_0, X_1) + W_2(K') = \\ = 2 - [e^{-X_1} + e^{-2(K_0 - X_1)}] + W_2(K'),$$

a last/latter term is located through the curve/graph of Fig. 3.27 upon the entrance into it with argument $K' = 0.75X_1 + 0.3(K_0 - X_1)$, where $K_0 = 2$.

Determining on only curved (see Fig. 3.28) maximum, let us find the final (no longer conditional) value of maximum income in all of five years:

$$W_{max} = \widetilde{W}_1(2) = 4,35$$

and the corresponding to it unconditional optimum control at the first step/pitch:

$$x_1 = 1,6.$$

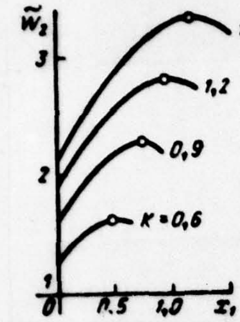


Fig. 3.26.

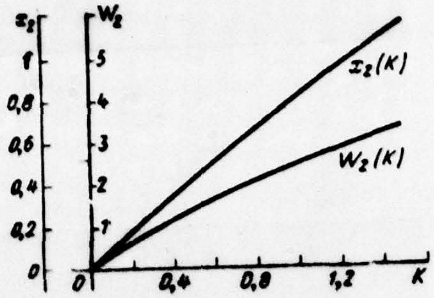


Fig. 3.27.

Page 152.

6. After the process of construction of conditional optimum controls and gains is finished, it is necessary to lead second stage of optimization, passing, step by step, control process from first step/pitch to latter on chain/network:

$$x_1 \to K_1^* \to x_2 \to K_2^* \to x_3 \to K_3^* \to x_4 \to K_4^* \to x_5 \to K_5^*.$$

Knowing x_1 = 1.6, we find the supply of means after the first step/pitch:

$$K_1^* = 0{,}75\, x_1 + 0{,}3\,(K_0 - x_1) = 1{,}32.$$

After entering with this value K_1 into graph x_2(K) in Fig. 3.27, we find optimum control on the second step/pitch:

$$x_2 = 1{,}02.$$

The residue/remainder of means after the second step/pitch will be:

$$K_2^* = 0{,}75\,x_2 + 0{,}3\,(K_1^* - x_2) = 0{,}86.$$

With this value K_2^* we enter in graph $x_3(K)$ (see Fig. 3.25) and we find optimum control on the third step/pitch

$$x_3 = 0{,}62.$$

Residue/remainder of means after the third step/pitch:

$$K_3^* = 0{,}75\,x_3 + 0{,}3\,(K_2^* - x_3) = 0{,}54.$$

Through the curve/graph of Fig. 3.23, we find optimum control on the fourth step/pitch

$$x_4 = 0{,}30.$$

Residue/remainder of means after the fourth step/pitch:

$$K_4^* = 0{,}75\,x_4 + 0{,}3\,(K_3^* - x_4) = 0{,}30.$$

With this value K_4^* we enter in graph $x_5(K)$ (see Fig. 3.21) and we find optimum control on the last/latter step/pitch $x_5 = 0$.

Thus, gliding/planning is finished: obtained optimum control, which indicates, how many means with their initial supply $K_0 = 2$ must be packed into branch I on years. This control will be:

$$x = (1{,}60;\ 1{,}02;\ 0{,}62;\ 0{,}30;\ 0).$$

Taking into account that the available means prior to the beginning of each year are known and equal to:

$$K_0 = 2;\quad K_1^* = 1{,}32;\quad K_2^* = 0{,}86;\quad K_3^* = 0{,}54;\quad K_4^* = 0{,}30,$$

we immediately find quantities of means, packed into branch II

through the years:

$$y = (0,40;\ 0,30;\ 0,24;\ 0,24;\ 0,30).$$

Thus, it is possible to formulate following recommendations regarding the insertion of means.

Page 153.

From the available in the beginning supply K_0 = 2 and the remaining means at the end of each year, it is necessary to pack on years in branch I and II following sums:

(1) Отрасли	(2) Год				
	1	2	3	4	5
I	1,60	1,02	0,62	0,30	0
II	0,40	0,30	0,24	0,24	0,30

Key: (1). Branches. (2). Year.

During this distribution of means in five years, will be obtained the maximum income, equal to

$$W_{max} = 4,35.$$

The residue/remainder of means at the end of the period will be equal to: 0.3•0.30 = 0.09.

Figures 3.29 depicts optimum trajectory in the phase space (each stage, except the first, is divided into half-steps).

From the examined example it is evident, how complex and tedious is step-by-step optimization "by hand", even for the most elementary problems (only two branches of productions; the simplest "functions of income" and the "function of expenditure"). Under any more complex conditions the development of optimum plan/layout of the method of dynamic programming is virtually impossible without the enlistment of high speed ETsVM.

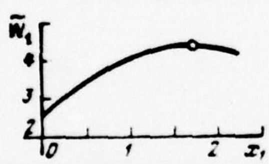


Fig. 3.28.

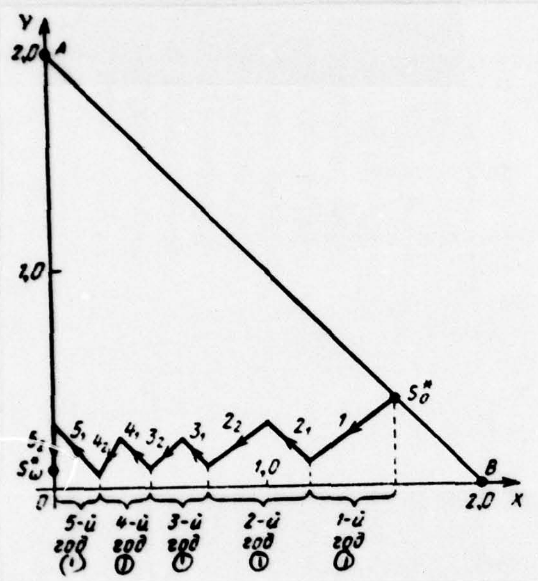


Fig. 3.29.

Key: (1). year.

Page 154.

6. Other problems of distributing the resource/lifetimes.

The problem of distributing the resource/lifetimes has many versions. Some of them comparatively differ little from the simplest problem, examined in §§ 4 and 5, others are so dissimilar to it on its verbal formulation, which until the next time is difficult to catch in them common/general/total features. Here and in following

paragraph we will give several examples of similar problems.

1. Distribution of resource/lifetimes in heterogeneous stages.

On the problem, examined in §4, stages were uniform in the sense that the "functions of the income" of f(X), of g(Y) and "the function of expenditure $\varphi(X), \psi(Y)$ were identical for all step/pitches. It can seem that they vary from one step/pitch to the next, namely for the i step/pitch they are equal to:

$$\left.\begin{matrix} f_i(X), & g_i(Y) \\ \varphi_i(X), & \psi_i(Y) \end{matrix}\right\} \quad (i=1,\ 2,\ \ldots,\ m).$$

In this case the standard set-up of the solution of problem barelys change. Basic functional equation takes the form

$$W_i(K)=\max_{0<X_i<K}\{f_i(X_i)+g_i(K-X_i)+ \\ +W_{i+1}(\varphi_i(X_i)+\psi_i(K-X_i))\}.$$

The condition of the optimization of the m step/pitch will be:

$$W_m(K)=\max_{0<X_m<K}\{f_m(X_m)+g_m(K-X_m)\},$$

a in all the remaining procedure of the construction of solution it will remain constant.

2. Problem of the redundancy of resource/lifetimes.

There is a total of one branch of production and certain supply of means K_0, which can be packed into production not wholly, but partially to reserve. If at the i step/pitch of production are

invested means X, then they give income $f_i(X)$ and are reduced to $\varphi_i(X)$.

It is required to rationally distribute the available and remaining means on m of step/pitches so that the total income for all m of step/pitches would be maximum.

Page 155.

It is not difficult to see that this problem is reduced to previous real/actually, the reserved means can be considered inserted into some fictitious second branch where they are not expended, but also do not give the income:

$$g_i(Y)=0; \quad \psi_i(Y)=Y \quad (i=1, \ldots, m).$$

Taking into account this condition the problem is solved in exactly the same way just as problem of distributing the resource/lifetimes in heterogeneous stages. The geometric interpretation of problem in phase space is shown on Fig. 3.30.

Let us consider the special case of the problem of the redundancy of resource/lifetimes, when in all stages

$$\varphi_i(X)=0 \quad (i=1, \ldots, m),$$

i.e. the inserted means are expend/consumed by pillar (Fig. 3.31). Since means are expended by pillar, then each horizontal trajectory phase reaches the very axis of ordinates.

Stated problem is reduced to finding of the maximum of function m of arguments $(X_1, X_2, ..., X_m)$:

$$W = \sum_{i=1}^{m} f_i(X_i),$$

where $X_1, X_2, ..., X_m$ are nonnegative and limited by the condition:

$$\sum_{i=1}^{m} X_i \leq K_0. \qquad (6.1)$$

If income $f_i(X)$ (as this logical to assume) represents by itself the nondecreasing function of the inserted means X, then inequality sign in formula (6.1) can be reject/thrown, since under these conditions to expend/consume means not to end is disadvantageous.

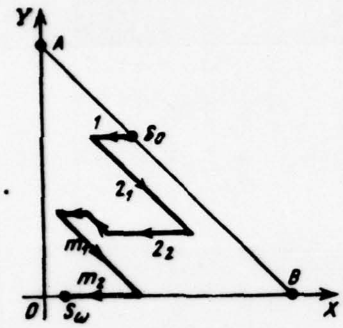


Fig. 3.30.

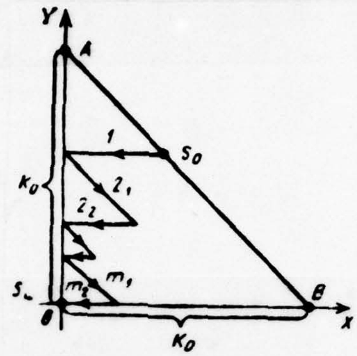


Fig. 3.31.

Page 156.

Let us note that some simplest problems of the redundancy of resource/lifetimes admit elementary solution also without the method of dynamic programming. To them belongs, for example, the simplest case when the "function of income" in all stages one and the same:

$$f_1(X)=f_2(X)= \ldots =f_m(X)=f(X)$$

and means are expend/consumed completely:

$$\varphi_1(X)=\varphi_2(X)= \ldots =\varphi_m(X)=0.$$

It is not difficult to ascertain that if the function of income is convex downward (Fig. 3.22), then it is more advantageous anything to put all means in some stage, and into the others not to pack. But

if the function of income is convex upward (Fig. 3.33), then the maximum of income is reached during the even distribution of means between the stages: $x_1 = x_2 = \ldots = x_m = K_0/m$.

3. Problem of distributing the rescurce/lifetimes between three and more by branches.

Let us assume that under conditions of problem of § 4 resource/lifetimes are distributed not between two branches (I and II), but between several: I, II, ..., (n): moreover for each (the j-th) branch are assigned: the "function of income" $f_i^{(j)}(X)$, expressing the income, yielded by means X, inserted on the i year into the j-th branch, and the "function of expenditure" $\varphi_i^{(j)}(X) < X$, showing, how much decrease means X, inserted on the i year into the j-th branch.

Problems it differs from that examined in the point/item of 1 this paragraph only by dimensionality (number of parameters, which determine the state of system). For example, for three branches I, II and III phase space is shown on Fig. 3.34. For the case of more than three branches geometric interpretaticn loses clarity, but the essence of problem remains the same. The state of system will be determined no longer by the pair of numbers X, Y, but of n by the numbers $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$,

designating insertions into each of the branches.

Fig. 3.32.

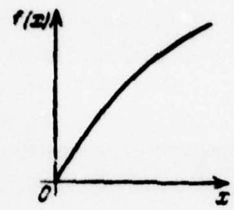


Fig. 3.33.

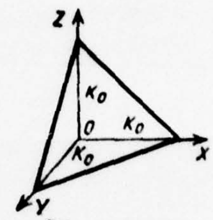


Fig. 3.34.

Page 157.

The process of distributing the means, as in the two-dimensional case can be divided into stages and produced first conditional optimization (from end at the beginning), and then - unconditional (from beginning toward the end). The state of the system prior to the beginning of each step/pitch will be as before characterized by the sum of the means, subject to distribution, i.e., by one number K.

Will more complex be matter with control. Control at the i step/pitch will consist of the isolation/liberation of means not of one branch, but of n of the branches:

$$X_i^{(1)},\ X_i^{(2)},\ \ldots,\ X_i^{(n)} = K - \sum_{i=1}^{n-1} X_i^{(i)}.$$

It is necessary to find the maximum of the function of several variables. With the number of branches n > 3 problems of optimization, it does become very bulky and without aid of EVM [ЭВМ - computer] scarcely it can be solved.

7. Distribution of resources with insertion of incomes into production.

Until now, in the problems of the distribution of resource/lifetimes, we examined the "income", yielded by enterprises, it is completed independent of the distributed means; it even could be expressed in other units (for example, resource/lifetimes - in man-hours, and income - in rubles). Now we will consider the case when income is packed into production (in full or in part). It goes without saying that for this income and means must be given to single (money) equivalent.

The problem of the distribution of resource/lifetimes with the insertion of incomes into production can be placed differently, depending on whether is packed the income in full or in part and which value is maximized.

Is given below series of problems, in each of which occurs the speech about the distribution of resource/lifetimes according to two branches of production with the insertion (full/total/complete or partial) of incomes into production, during different objective functions.

1. Income is packed into production completely, is maximized sum of all means (basic plus income) after m stage.

In this case gain W represents by itself the sum of all means, which were preserved in both branches after the completion of last/latter stage, plus the income, given by both branches in last/latter stage. This entire gain is acquired only on one, last/latter stage, but it represents by itself a special case of the additive index of efficiency for which

$$W=\sum_{i=1}^{m} w_i,$$

if we consider that the gains in all stages, except the latter, are equal to zero

$$w_1=w_2=\ldots=w_{m-1}=0; \quad w_m=W.$$

Page 158.

Since all means (and basic and income) are packed into production on equal basis/bases, there is no necessity to examine separately the "functions of income" and of "function expend", and it suffices to introduce for each branch only according to one function: for branch I - function $F_i(X)$, showing, how many means (including income) it will be obtained at the end of the i step/pitch in branch I with the insertion of it of means X in the beginning of this

step/pitch. Analogous functicn for branch II will be $G_i(Y)$. Let us name the functions

$$F_i(X), \quad G_i(Y)$$

the "functions of a change in the means" in the i stage. Let us note that is generally possible any of the relationship/ratios:

$$F_i(X) > X; \quad F_i(X) < X; \quad F_i(X) = X$$

(it is analogous for $G_i(Y)$).

Let us consider the phase space, ccrresponding to this problem (Fig. 3.35). This space will be no longer triangle AOB (as in problems without the insertion of incomes), but entire first quadrant XOY (means can not only be reduced, but alsc increase). Trajectory consists as before of a series of the compcnent/links, being decomposed into half-sections; the first half-section (for all stages, except the first) represents redistribution of the means (point S moves in parallel $\overline{AB}$), the second - expenditure and acquisition of the means (pcint S can mcve in any direction). Unlike previously examined problems, here incame yields only one, the latter component/link that in Fig. 3.35 is isclated by heavy arrow.

In this case the value of index W is directly evident on drawing - this sum of abscissa and crdinate of the point S_ω, of the representing final state system. Problem of the optimum control: to deduce point S_ω on straight line $A_\omega B_\omega$, parallel AB and outermost from

the origin of coordinates. The value of gain for any trajectory in phase space represents by itself each of the segments, intercept/detached by straight line $A_\omega B_\omega$ on the coordinate axes.

Let us construct the set-up of the solution of this problem by the method of dynamic programming, without detailed verbal explanations (throughout the specimen/sample of the previous problems). On function $F_i(X)$ and $G_i(Y)$ thus far let us set no limitations.

Gain at all step/pitches, except the latter, is equal to zero; therefore we will not it record/write.

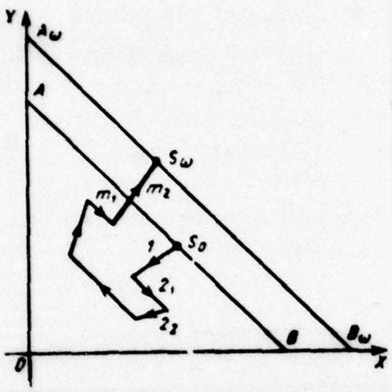


Fig. 3.35.

Page 159.

At last/latter step/pitch it is expressed by the formula:

$$w_m(K, X_m) = F_m(X_m) + G_m(K - X_m), \quad (7.1)$$

where K - the means with which we approached the last/latter step/pitch.

The fundamental functional equation of dynamic programming will be:

$$W_i(K) = \max_{0 < X_i < K} \{W_{i+1}(F_i(X_i) + G_i(K - X_i))\}, \quad (7.2)$$

where K - the means with which we approached the i step/pitch.

At last/latter step/pitch we obtain the conditional optimum gain, equal to

$$W_m(K) = \max_{0 < X_m < K} \{F_m(X_m) + G_m(K - X_m)\}, \quad (7.3)$$

and condition the optimum control at which this gain is reached: $x_m(K)$.

Further, through formula (7.2) we find everything conditional gain and conditional optimum controls on all step/pitches, beginning with the latter, after which process it passes in forward direction they are determined unconditional optimum controls at each step/pitch.

Is such the set-up of the solution of problem by the method of dynamic programming with any form of the function of a change in means $F_i(X), G_i(Y)$. However, if we on these functions superimpose some (very natural) limitation, set-up can be highly simplified.

Let us assume that all the functions $F_i(X), G_i(Y)$ represent by themselves the nondecreasing functions of their arguments, i.e., with an increase in the quantity of inserted means, the sum of income and remaining means toward the end of the stage cannot decrease.

Let us show that in this case conditional optimum gain there is the nondecreasing function from the issue of each of the previous step/pitches, i.e. from the sum of means of its end.

Actually, let the issue of some, let us say, that (i-1) step (sum of means of its end) is equal to K_{i-1}. Let us consider optimum gain under this condition as function K_{i-1}. Since gain is acquired only at last/latter step/pitch, then it is unimportant, to examine this gain for all the step/pitches, either only for last/latter step/pitch, or for all the step/pitches beginning with the i-th. Let us select the latter: let us consider the optimum gain W for all the step/pitches by beginning with the i-th as function K_{i-1}:

$$W_i(K_{i-1}). \qquad (7.4)$$

It is necessary to demonstrate that this function - nondecreasing. Proof let us conduct the method of full/total/complete induction, but not from one i to i + 1, but, on the contrary, from one i + 1 to next. Let us assume that the proven property is correct for i + 1, i.e.

$$W_{i+1}(K_i) \qquad (7.5)$$

there is the nondecreasing function of its argument K_i (sum of means at the end of the i step/pitch). Let us demonstrate that then nondecreasing function it will be and (7.4).

Page 160.

It is real/actual, according to equation (7.2) (where K_{i-1} is marked simply K) function $W_i(K_{i-1})$ represents by itself the maximum of the expression

$$W_{i+1}(F_i(X_i)+G_i(K_{i-1}-X_i)) \qquad (7.6)$$

Let us show that (7.6) there is the nondecreasing function from K_{i-1}; then will be it is clear that as its maximum value $W_i(K_{i-1})$ with an increase K_{i-1} decrease cannot.

Let us fix some value K_{i-1}. Let for this value expression (7.6) reach maximum in X_i, equal to $W_i(K_{i-1})$, during the specific control x_i.

Let us give now to value K_{i-1} certain positive increase ΔK_{i-1}. For us was formed certain surplus of means, which we can put additionally either into branch I or into branch II, or into both immediately. Since function $F_i(X), G_i(Y)$ nondecreasing, the from this "addition" of means each term under the sign of function (7.6) can only be increased, and also, therefore, their sum it can only be increased, but not shape less. What in this case will stop with function (7.6)? According to our assumption, function W_{i+1} — nondecreasing, which means, that with an increase K_{i-1} expression (7.6) decrease cannot. Thus, transition from i + 1 to i is proven.

Let us show now that our property is correct for a last/latter step/pitch (i + 1 = m). This is proven simply. On formula (7.3) the gain at the m step/pitch during optimum control represents by itself the maximum of the expression

$$F_m(X_m)+G_m(K_{m-1}-X_m)$$

and, it is logical, it is nondecreasing function from K_{m-1} (this recently it was proved for any i, and also, therefore, for i = m). Thus, $W_m(K_{m-1})$ there is nondecreasing function K_{m-1}, a that means that according to the principle of full/total/complete induction, and any of gains $W_i(K_{i-1})$ — nondecreasing function, which it was required to prove.

From that demonstrated escape/ensue very simple recommendations regarding optimum control. It is real/actual, if final optimum gain there is nondecreasing function from the common/general/total sum of means, realized on the issue of each step/pitch, then optimum control lies in the fact that, as a result of each step/pitch obtaining the of maximum value of this sum means. That means that control of each separate step/pitch can be chosen on the basis of the interests of this separate step/pitch, without taking into account the others.

This special feature/peculiarity of stated problem leads to the fact that the process of gliding/planning strongly is simplified. There is no already necessity for the complex procedure of the determination of conditional optimum gains and conditional optimum controls - for each step/pitch, beginning with the first, immediately it is located unconditional optimum control. At the first step/pitch it is necessary to select the control x_1, during which it is converted into maximum K_1 - sum of means after the first step/pitch:

$$K_1^* = \max_{0 \leq X_1 \leq K_0} \{F_1(X_1) + G_1(K_0 - X_1)\}.$$

Page 161.

On the second - the control during which is converted into maximum **value** $F_2(X_2) + G_2(K_1^* - X_2)$:

$$K_2^* = \max_{0 \leqslant X_2 \leqslant K_1^*} \{F_2(X_2) + G_2(K_1^* - X)\},$$

and so **forth to** K_{m-1}^*. **Maximum gain** at the m step/pitch will be equal to:

$$W_m = \max_{0 \leqslant X_m \leqslant K_{m-1}^*} \{F_m(X_m) + G_m(K_{m-1}^* - X_m)\}.$$

Thus, during nondecreasing functions $F_i(X), G_i(Y)$ **stated** problem of distributing the resource/lifetimes **is only outwardly** similar not the problem of dynamic programming, but actually - is much simpler it.

The similar degenerate problems of the dynamic programming where the optimum control consists of the simple optimization of each step/pitch, frequently are encountered in practice. If, having focused attention on this special feature/peculiarity, to solve them all the same method of dynamic programming, solution, it goes without saying, it will be accurate, but will require many times of more time, than if we immediately take into account their degeneracy.

2. Income is packed into production completely in all stages,

except latter; we are maximized income at last/latter step/pitch.

Problem differs from that examined above by the fact that is maximized not the sum of the remaining means plus income at last/latter step/pitch, but only one income at last/latter step/pitch, regardless of the fact, how many means were preserved from initially inserted.

In order to separate/liberate the sum of the remaining means from income, it is necessary for a last/latter step/pitch to assign not the function of a change in the means, but separately of the "function of income" $f_m(X), g_m(Y)$ and the "function of expenditure" $\varphi_m(X), \psi_m(Y)$.

It is easy to ascertain that the problem so placed, is reduced to previous. It is real/actual, set/assuming at the last/latter step/pitch

$$F_m(X)=f_m(X); \quad G_m(Y)=g_m(Y),$$

we obtain conditions p. 1. It is logical that if all the functions $F_i(X), G_i(Y)$ (i = 1, ..., m) - nondecreasing, this problem, as previous, will be degenerated.

3. Income is packed into production not completely, but some part of it is dropped from the roll; is maximized full/total/complete

deducted income in all stages plus residue/remainder of means after m stage.

Page 162.

For the solution of this problem, must be assigned to the "function of income":

$$f_i(X), \quad g_i(Y) \quad (i=1, \ldots, m),$$

the "functions of expenditure":

$$\varphi_i(X) \leqslant X; \quad \psi_i(Y) \leqslant Y \quad (i=1, \ldots, m),$$

and still, additionally, the "function of deductions":

$$r_i(D) \leqslant D \quad (i=1, \ldots, m),$$

showing, what part of income D, obtained at the i step/pitch, is not packed into production at following (i + 1) step, but it is dropped from the roll.

Let us plan the set-up of the solution of problem by the method of dynamic programming. The state of the system prior to the beginning of the i step/pitch let us characterize a quantity of means K, which are subject to distribution; it is obtained from the issue of the previous step/pitch via the deduction of the specific fraction of income.

Gain at the i step/pitch will be

$$w_i(K_i X_i) = r_i(f_i(X_i) + g_i(K - X_i)).$$

Control X_i at the i step/pitch (insertion of means X_i into branch I, and remaining means - into branch II) translates system from state K into the new state:

$$K'=\varphi_i(X_i)+\psi_i(K-X_i)+f_i(X_i)+$$
$$+g_i(K-X_i)-r_i(f_i(X_i)+g_i(K-X_i)).$$

Fundamental functional equation:

$$W_i(K)=\max_{0\leqslant X_i\leqslant K}\{r_i(f_i(X_i)+g_i(K-X_i))+$$
$$+W_{i+1}(\varphi_i(X_i)+\psi_i(K-X_i)+f_i(X_i)+$$
$$+g_i(K-X_i)+r_i(f_i(X_i)+g_i(K-X_i)))\}.$$

Conditional optimum gain at the m step/pitch:

$$W_m(K)=\max_{0\leqslant X_m\leqslant K}\{f_m(X_m)+g_m(K-X_m)+\varphi_m(X_m)+\psi_m(K-X_m)\}.$$

In the remaining set-up of dynamic programming remains the same, as before for the nondegenerate problems of distributing the resource/lifetimes.

We recommend to reader as an exercise to sketch the set-ups of the solution of the following problems of distributing the resource/lifetimes.

4. Income is packed into production not completely, but partially; maximized only full/total/complete deducted income for all

m of step/pitches, without account of the remaining means.

5. Income is packed into production not completely, but partially; is maximized total quantity of means (basic plus income) after m step/pitch, without account previously deducted sums.

Will not be any of these problems under some conditions degenerated?

Page 163.

8. Solution of the problem of dynamic programming taking into account the prehistory of process.

All problems of the dynamic programming which we, until now, examined, differed in terms of the following special feature/peculiarity: "income" w_i at each i step/pitch and the maximum income W beginning with the i step/pitch and they further depended only on state S of system S before this, i step/pitch and on the used control U_i, but they did not depend on how (in what way) system arrived into state S, i.e., as a result of which controls when and as this occurred. By other owls, the problem of the optimization of control at each (i-th) step/pitch we solved taking into account the present state S, but without the account to the

prehistory of process.

For example, solving the problem of distrituting the resource/lifetimes between by two (or it is more) branches of production, we as the characteristic of the state of the system before each step/pitch took one value - available available means K; to us was in no way matters to that, when and as system it arrived into this state, i.e., as were distributed means between branches over all previous stages. Was importantly only quantity of means K, with which we arrived at next step/pitch.

In many problems of dynamic programming, this "independence from prehistory" does not occur. For example, income at the i step/pitch can depend not only on the quantity of means, inserted into each branch at this step/pitch, but even on which means and at which step/pitches were packed into it earlier.

Theoretically always it is possible to take into account the prehistory of process with the help of the following method: to include/connect in the number of phase coordinates, characterizing state S of system s before this step/pitch, all those parameters from the past, on which depends future.

For example, if income at the i step/pitch depends not only on

the inserted means X_i, but also on previously inserted means Z, it is possible to characterize the state of the system before the i step/pitch by the simply not available available supply of means K, but set (K, Z), where Z - previously inserted means.

If is essential not only common/general/total sum previously inserted means, but also when precisely and how many means were packed - in principle it is possible "to enrich" state S and by these information from the past. Thus, theoretically always it is possible to introduce into the number of parameters, which characterize the state of system at present torque/moment, as much as desired the parameters from "the past". However, in practice this "enrichment" of phase space rapidly leads to boundless compound circuit of dynamic programming, with so complex that the method itself ceases to be suitable. Indeed the main idea of the dynamic programming: "instead of one time solving of complex problem, many times solving comparatively simple" ceases itself to justify, if "simple" problem ceases to be "idle time".

Page 164.

Therefore the attempts to solve by the method of the dynamic programming of problem with complex effect "prehistory" usually to nothing good do not lead.

However, if the effect of "prehistory" can be taken into account with the help of the small number of parameters (one, two, three), sometimes to construct a comparatively simple set-up of dynamic programming and it is possible to solve the problem of optimization.

As an example of problem "with prehistory" let us consider the problem of the maintenance of technology.

Problem is placed as follows.

There is the technical equipment/device S, exploited during m of years.

Operating costs depend on the following factors:

- from the "age" of equipment/device t, i.e., quantity of years, its past from input time into operation;

- from a quantity of maintenance k, produced to torque/moment t;

- from a quantity of years τ, of past from time last/latter maintenance [1].

FOOTNOTE [1]. Strictly speaking, operating costs depend not only on time τ, passed after last/latter repair, but also on periods previous k of repairs; but this dependence is weak, and it it is possible not to consider. ENDFOOTNOTE.

Let us assume that the maintenance is produced (if it is produced) instantly, also, in the beginning of year. It is logical to assume that the expenditures on this repair (cost/value of repair) depend on the same arguments t, k and τ that and operating costs.

We wish so to distribute maintenance on years, in order to the sum of overall expenditures (operating costs plus expenditure/consumptions to repair, if it was produced) they would reach the minimum.

Stated problem can be solved by the method of dynamic programming, if we characterize the state of system (technical equipment/device S) at the beginning of each step/pitch by three phase coordinates; t - the "age" of system, k - by quantity of repairs in the past and τ - by time, past from the torque/moment of last/latter repair.

In order to solve the problem of the optimization of control, it is necessary to assign both the operating costs and the expenditure/consumptions to repair in function from these phase coordinates.

Let us introduce following designations.

$Э_0(t)$ — the cost/value of the operation of equipment/device for the year, which begins at torque/moment t, if to torque/moment t of no repair it was produced;

$Э_1(t, \tau)$ -- the cost/value of the operation of equipment/device for the year, beginning at torque/moment t, if to torque/moment t was produced one repair, and from the time of this repair it passed τ years;

and generally

$Э_k(t, \tau)$ — cost/value of the operation of equipment/device for one year, beginning at torque/moment t, if to torque/moment t was produced k of repairs, and from the time of the latter of them passed τ years.

Page 165.

$R_0(t)$ - the cost/value of the repair, produced at torque/moment t, if to torque/moment t of no repair it was produced;

$R_1(t, \tau)$ - the cost/value of the repair, produced at torque/moment t, if to torque/moment t was produced one repair, and from the time of this repair it passed τ years;

and generally

$R_k(t, \tau)$ — cost/value of the repair, produced at torque/moment t, if to torque/moment t was produced k of repairs, and from the time of the latter of them passed τ years.

Let us represent the state of the technical equipment/device S as point S in phase space; along one axis we will plot/deposit the "age" of equipment/device - time t, on another - the time τ, past from the torque/moment of last/latter repair, on the third - quantity of repairs k (Fig. 3.36). Since under all conditions $\tau < t$ and $k < t$, then all the possible states of system will be represented as points within trihedral angle OAtB. If to torque/moment t of repair it was not, point S is located on axis Ot; if was one repair - point S it is located in plane KO'L, parallel tOτ and by that distant behind it up

to distance of 1, and sc forth.

In order not to use three-dimensiocal/space picture, "it is stratified" phase space on several parts which we will designate:

(0), (1), (2), ..., (*k*), ...

Part (0) of phase space represents by itself simply axis 0t; part (1) - triangle on plane K0'L, part (2) - triangle on the plane, parallel t0$_T$ and lying of it at a distance of 2 and so forth. With an increase in the number of space the size/dimensions of triangles always are reduced. The parts of phase space (0), (1), (2), ..., (k), ... are shown on Fig. to 3.37.

Prior to the beginning of each year of us, exists a selection between two controls:

U^0 - not to make repair (to continue to exploit equipment/device S).

U^1 - to do a repair (and after it to continue to exploit equipment/device).

Let us look, which transferring in phase space experiences point S under the action of each control.

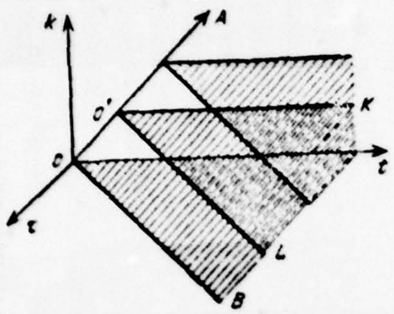


Fig. 3.36.

Page 166.

Let point S is locate in space (0) - on axis Ot at point with coordinate t (see Fig. 3.37). Under the effect of control U^0 (to continue to exploit) it for year will move into point with abscissa t + 1 on the same axis.

Under the effect of control U^1 (to do a repair) it will move into point S' in space (1) with coordinates (t + 1, 1). The second coordinate τ = 1, since repair is produced in the beginning of year, i.e., for year to end of next step/pitch.

Now point S occupies some attitude (1). Control U^0 (to continue to exploit) will lead to the fact that both t and τ for one step/pitch will be increased on one unit, i.e., point S will move upward and to the right (in parallel the hypotenuse of triangle) into point with coordinates (t + 1, τ + 1), if previous coordinates were (t, τ).

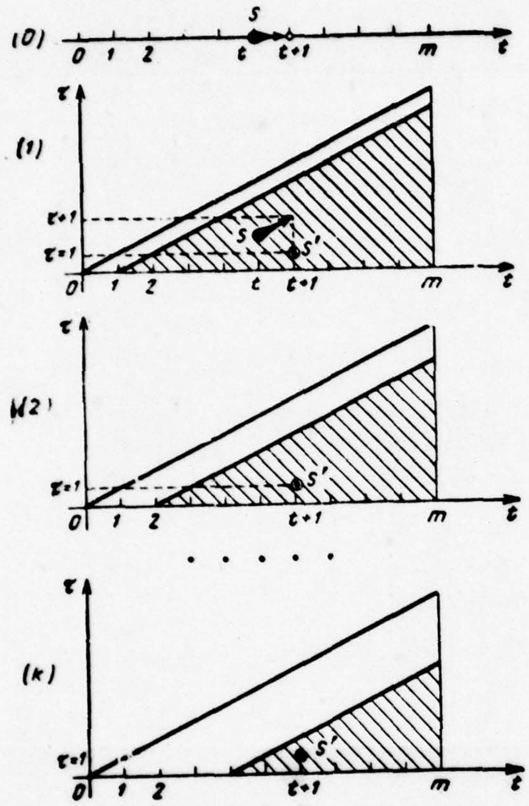


Fig. 3.37.

Page 167.

But if we use control U' (let us do a repair), point will move into space (2), into point S' with coordinates (t + 1, 1).

Generally, if point S is located in space (k) (k ⩾ 1), then control U⁰ moves it to one step/pitch to the right and upward, of

point with coordinates (t, τ) of point with coordinates $(t + 1, \tau + 1)$, but control U^1 - of the following in order space $(k + 1)$, of point with coordinates $(t + 1, 1)$.

Let us register the rules of the transition of point S in phase space under the effect of controls U^0 and U^1 in the form of the "table of transformation" (see Table 8.1, the first of five columns).

Thus, to us it is clear, as is moved the point in phase space under the effect of any control, i.e., we know the function

$$S' = \varphi(S, U),$$

according to which it varies the state of system under the effect of the used control U $(U = U^0, U^1)$.

Now let us look, to which "gain" - to the expenditure/consumption of w_i at this step/pitch will bring each control. If we will use control U^0, then at this step/pitch we will have only operating costs; if control U^1 - expenditure/consumptions to repair plus operational to the nearest year, but others, than if repair was not. Let us register these expenditure/consumptions in the same Table 8.1 in the form of additional column.

Using this table, we can now for any state of system S and any control (U^0 or U^1), used at the any moment t, to find:

- where will move point S under control effect;

- to which expenditure of resources this will lead.

After this table it is comprised, it is already not difficult to organize the very procedure of optimization.

Table 8.1.

(1) Исходное положение		(5) Управление	(2) Новое состояние		(6) Расход на данном шаге, начинающемся в момент t
(3) пространство	(4) координаты		(3) пространство	(4) координаты	
(0)	(t)	U^0 U^1	(0) (1)	$(t+1)$ $(t+1, 1)$	$Э_0(t)$ $R_0(t)+Э_1(t, 0)$
(1)	(t, τ)	U^0 U^1	(1) (2)	$(t+1, \tau+1)$ $(t+1, 1)$	$Э_1(t, \tau)$ $R_1(t, \tau)+Э_2(t, 0)$
. .	. .	. . . . .	. . .	. . .	. . . . . .
(k)	(t, τ)	U^0 U^1	(k) $(k+1)$	$(t+1, \tau+1)$ $(t+1, 1)$	$Э_k(t, \tau)$ $R_k(t, \tau)+Э_{k+1}(t, 0)$
. .	. .	. . . . .	. . .	. . .	. . . . . .

Key: (1). Initial position. (2). New state. (3). space. (4). coordinate. (5). Control. (6). Expenditure at this step/pitch, which begins at torque/moment t.

Page 168.

We will begin, as ever, from last/latter step/pitch, let us sort out all the possible states of the system before this step/pitch and for each of them will find conditional optimum control (U^0 or U^1) on the m step/pitch and conditional optimum gain (minimum expenditure) on last/latter step/pitch. Further let us optimize (m-1) step/pitches, so as to it, in conjunction with the already optimized m-th, would give minimum expenditure, etc.

Let us demonstrate this methodology on concrete/specific/actual example.

Example: a section of railway line is exploited during m = of 6 years. Operating costs for one year, which begins at torque/moment t, (in arbitrary units) are expressed by the functions: $Э_0(t)$, $Э_1(t, \tau)$, $Э_2(t, \tau)$, $Э_3(t, \tau)$, $Э_4(t, \tau)$, $Э_5(t, \tau)$ (index - quantity of previous repairs, τ - quantity of years, past from the last/latter repair of way). The cost/value of repair is assigned by functions $R_0(t)$, $R_1(t, \tau)$, $R_2(t, \tau)$, $R_3(t, \tau)$, $R_4(t, \tau)$. The values of functions Э and R are given in Tables 8.2 and 8.3.

Table 8.2.

Функция (1)	t	0		2	3	4	5
$Э_0(t)$		1,9	2,5	3,1	4,0	5,1	6,6
$Э_1(t,\tau)$	τ \ t	0	1	2	3	4	5
	0	—	2,2	2,4	3,8	5,0	6,3
	1	—	—	2,5	3,9	5,1	6,4
	2	—	—	—	4,0	5,1	6,5
	3	—	—	—	—	5,1	6,6
	4	—	—	—	—	—	6,6
$Э_2(t,\tau)$	τ \ t	0	1	2	3	4	5
	0	—	—	2,3	3,7	4,8	5,5
	1	—	—	—	3,8	4,9	5,7
	2	—	—	—	—	5,0	6,0
	3	—	—	—	—	—	6,2
$Э_3(t,\tau)$	τ \ t	0	1	2	3	4	5
	0	—	—	—	2,8	3,9	4,5
	1	—	—	—	—	4,0	4,7
	2	—	—	—	—	—	5,5
$Э_4(t,\tau)$	τ \ t	0	1	2	3	4	5
	0	—	—	—	—	3,5	4,2
	1	—	—	—	—	—	4,5
$Э_5(t,\tau)$	τ \ t	0	1	2	3	4	5
	0	—	—	—	—	—	3,8

Key: (1). Function.

Page 169.

Table 8.3.

(1) Функция	τ \ t	1	2	3	4	5
$R_0(t)$	—	1,2	1,4	1,8	2,3	3,0
$R_1(t, \tau)$	1	—	1,2	1,4	1,9	2,4
	2	—	—	1,5	2,0	2,5
	3	—	—	—	2,1	2,6
	4	—	—	—	—	2,9
$R_2(t, \tau)$	τ \ t	1	2	3	4	5
	1	—	—	1,2	1,5	2,0
	2	—	—	—	1,6	2,1
	3	—	—	—	—	2,3
$R_3(t, \tau)$	τ \ t	1	2	3	4	5
	1	—	—	—	0,8	1,1
	2	—	—	—	—	1,4
$R_4(t, \tau)$	τ \ t	1	2	3	4	5
	1	—	—	—	—	1,0

Key: (1). Function.

Table 8.4.

(1) Условная оптимизация шестого шага					
(2) Пространство	(3) Состояние системы (координаты t, τ)	(4) Расход при управлении U^0	(4) Расход при управлении U^1	(5) Оптимальное управление	(6) Минимальный расход
(0)	$t=5$	6,6	3,0+6,3=9,3	U^0	6,6
(1)	(5,1) (5,2) (5,3) (5,4)	6,4 6,5 6,6 6,6	2,4+5,5=7,9 2,5+5,5=8,0 2,6+5,5=8,1 2,9+5,5=8,4	U^0 U^0 U^0 U^0	6,4 6,5 6,6 6,6
(2)	(5,1) (5,2) (5,3)	5,7 6,0 6,2	2,0+4,5=6,5 2,1+4,5=6,6 2,3+4,5=6,8	U^0 U^0 U^0	5,7 6,0 6,2
(3)	(5,1) (5,2)	4,7 5,5	1,1+4,2=5,3 1,4+4,2=5,6	U^0 U^0	4,7 5,5
(4)	(5,1)	4,5	1,0+3,8=4,8	U^0	4,5

Key: (1). Conditional optimization of the sixth step/pitch. (2). Space. (3). State of system (coordinates t, τ). (4). Expenditure during control. (5). Optimum control. (6). Minimum expenditure.

Page 170.

Solution. Using the tables of functions 8.2 and 8.3 and the table of transformation 8.1, let us develop the process of dynamic programming. As ever, let us begin from the optimization of last/latter (the sixth) step/pitch.

All the possible states of system S before this step/pitch will be represented as points a by abscissa t = 5 in spaces (0), (1), (2),

(3), (4) (see Fig. 3.38). For the sixth (latter) step/pitch optimum will be the control (U^0 or U^1), during which the expenditure at last/latter step/pitch is minimal. Expenditures let us compute according to last/latter column Table 8.1. In Fig. 3.38, besides the state of system, we will designate even the optimum control: U^0 will be designated by the arrow/pointer, directed to the right (in space (0)) and to the right and upward (in remaining spaces). Control U^1, which removes point from this section of the phase space and which translates into following in order part, let us represent as the arrow/pointer, directed to the right and downward. For each point within small circle let us record/write the minimum expenditure at all remaining step/pitches, which corresponds to this state of system (conditional optimum gain).

Table 8.5.

(1) Условная оптимизация пятого шага					
(2) Пространство	(3) Состояние системы (координаты t, τ)	(4) Расход при управлении U^0	(4) Расход при управлении U^1	(5) Оптимальное управление	(6) Минимальный расход
(0)	$t=4$	5,1+6,6=11,7	2,3+5,0+6,4=13,7	U^0	11,7
(1)	(4,1) (4,2) (4,3)	5,1+6,5=11,6 5,1+6,6=11,7 5,1+6,6=11,7	1,9+4,8+5,7=12,4 2,0+4,8+6,0=12,8 2,1+4,8+6,2=13,1	U^0 U^0 U^0	11,6 11,7 11,7
(2)	(4,1) (4,2)	4,9+6,0=10,9 5,0+6,2=11,2	1,5+3,9+4,7=10,1 1,6+3,9+4,7=10,2	U^1 U^1	10,1 10,2
(3)	(4,1)	4,0+5,5=9,5	0,8+3,5+4,5=8,8	U^1	8,8
(7) Условная оптимизация четвертого шага					
(0)	$t=3$	4,0+11,7=15,7	1,8+3,8+11,6=17,2	U^0	15,7
(1)	(3,1) (3,2)	3,9+11,7=15,6 4,0+11,7=15,7	1,4+3,7+10,1=15,2 1,5+3,7+10,1=15,3	U^1 U^1	15,2 15,3
(2)	(3,1)	3,8+10,2=14,0	1,2+2,8+8,8=12,8	U^1	12,8
(8) Условная оптимизация третьего шага					
(0)	$t=2$	3,1+15,7=18,8	1,4+2,4+15,2=19,0	U^0	18,8
(1)	(2,1)	2,5+15,3=17,8	1,2+2,3+12,8=16,3	U^1	16,3
(9) Условная оптимизация второго шага					
(0)	$t=1$	2,5+18,8=21,3	1,2+2,2+16,3=19,7	U^1	19,7
(10) Оптимизация первого шага					
(0)	$t=0$	2,0+19,7=21,7	—	U^0	21,7

Key: (1). Conditional optimization of the fifth step/pitch. (2). Space. (3). State of system (coordinates t, τ). (4). Expenditure during control. (5). Optimum control. (6). Minimum expenditure. (7). Conditional optimization of fourth step/pitch. (8). Conditional optimization of third step/pitch. (9). Conditional optimization of second step/pitch. (10). Optimization of first step/pitch.

Page 171.

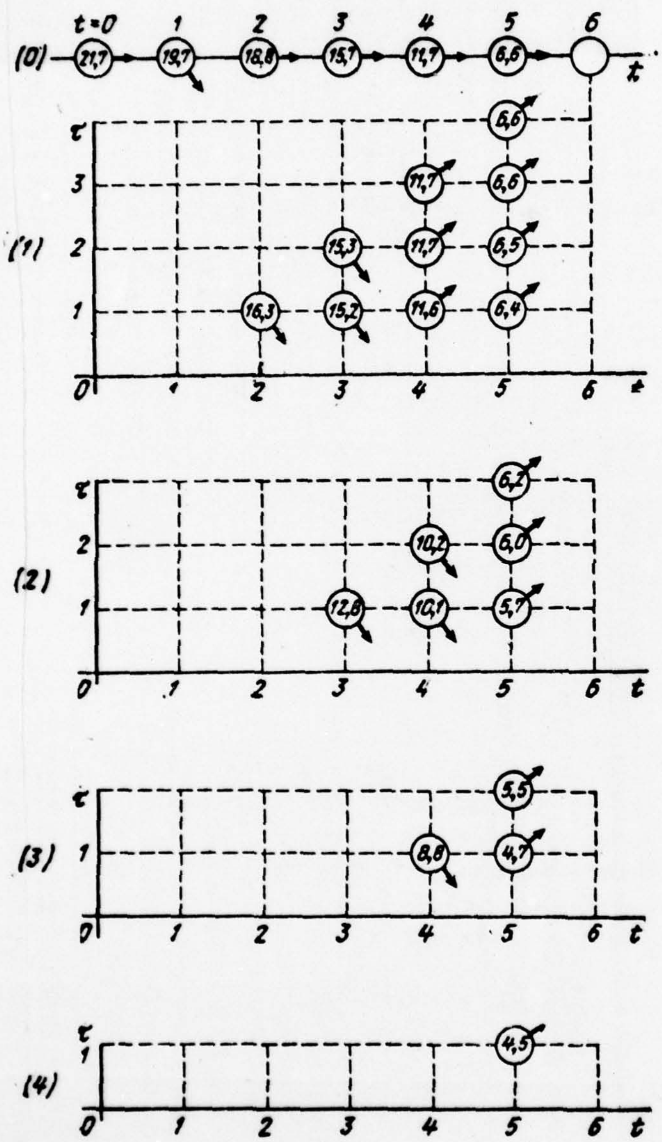


Fig. 3.38.

Page 172.

The calculations, connected with optimization, let us take shape in the form of the tables (see Tables 8.4 and 8.5 on page 169, 170).

Thus, optimization is finished. It will lead us to following conclusions.

Minimum expenditure/consumption is equal to 21.7. Is reached it at the following optimum control:

$$u = (U^0, U^1, U^1, U^1, U^1, U^0),$$

i.e.:

- on the first year section is exploited without repair;

- in the beginning of the second, third, fourth and fifth years is produced repair;

- on sixth year section is exploited without repair.

In this case, the expenditure/consumptions reach the minimum, equal to 21.7 arbitrary units [1].

FOOTNOTE [1]. During the analysis of this example, it is necessary to keep in mind that the numerical data are selected from methodical considerations and nothing in common with validity have. ENDFOOTNOTE.

9. Problems of dynamic programming, not connected with time.

Until now, we examine only such problems of the dynamic programming where the planned/glide operation is developed in time and falls into a series of the step/pitches (stages), following after each other in natural, time/temporary order - from the first step/pitch toward the latter. Generally, this not is compulsory: breakdown into steps or "staging" in the problems of dynamic programming can be produced not on time, but according to any other sign/criterion, for example, according to the reference number of one or the other object.

As an example let us consider following task.

Let there be the group of the enterprises

$$\Pi_1, \Pi_2, \ldots, \Pi_m, \tag{9.1}$$

which issue one and the same production. We have available - some

supply of means K_0, which we can put in the group of enterprises (9.1) in order to produce over plan/layout maximum output.

Let us assume that each enterprise can master only limited quantity of means, and

$$k_1, k_2, \ldots, k_m \qquad (9.2)$$

represent the maximum sums, which can master respectively enterprises (9.1). If in enterprise Π_i are invested means X, it will give $\varphi_i(X)$ unity of further (above-plan) production.

It is required so to distribute the available means between enterprises so that the total volume W of further production would be maximum.

Page 173.

Control of means lies in the fact that to enterprises are selected respectively the means:

$$X_1, X_2, \ldots, X_m,$$

not exceeding in the sum of available capital K_0:

$$\sum_{i=1}^{m} X_i < K_0,$$

it is required to find the optimum control, by which

$$W = \sum_{i=1}^{m} w_i = \max,$$

where w_i - further production of the i enterprise.

Stated problem is easily solved by the method of dynamic programming; the "stage" of the process of distributing the means is the isolation/liberation of means to the i enterprise.

Let us label stages (step/pitches) by way of numbers of enterprises (i.e. in arbitrary order). Let us assume that the means to enterprises $\Pi_1, \ldots, \Pi_{m-1}$ are already isolated, and at last/latter, m step/pitch we arrived with some supply of means K.

It is obvious, optimum control on last/latter step/pitch lies in the fact that, isolating m enterprise of all the remaining means K, if they do not exceed k_m, and a maximally possible quantity of means k_m, if $K \geqslant k_m$. Thus, conditional optimum control at the last/latter step/pitch:

$$x_m(K) = \begin{cases} K \text{ when } K \leqslant k_m, \\ k_m \text{ when } K > k_m. \end{cases}$$

During this control maximum income at last/latter step/pitch will be

$$W_m(K) = w_m(K) = \varphi_m(x_m(K)). \qquad (9.3)$$

Let us pass to gliding/planning of penultimate step/pitch - to the isclation/liberaticn cf means to (m - 1) enterprise. After m - 2 step/pitches available will remain means K. We must select this control

$$x_{m-1} \leqslant k_{m-1},$$

with which income at (m - 1) step/pitch plus the already optimized income on the latter is converted into the maximum:

$$W_{m-1}(K) = \max_{0 \leqslant x_{m-1} \leqslant K_{m-1}} \{\varphi_{m-1}(X_{m-1}) + W_m(K - X_{m-1})\}, \qquad (9.4)$$

and so forth.

The fundamental functional equaticn of dynamic programming will be

$$W_i(K) = \max_{0 \leqslant X_i \leqslant k_i} \{\varphi_i(X_i) + W_{i+1}(K - X_i)\}, \qquad (9.5)$$

a entire/all procedure of conditional and unconditional optimization in no way differs from that problem of the distribution of resource/lifetimes according to heterogeneous stages with redundancy, which we examined above, in §6.

Page 174.

Thus, the method of dynamic programming, which initially was presented to us as specific method of the optimization of the processes, which develop in time, has much wider field of application/uses.

Example. One must plan multistage space vehicle within the limits of the specific launching weight G. Cosmonaut's cab has the assigned weight g_K. It is assumed that the rocket will have m of step/stages. Launching weight of rocket is composed of the weights of all stages of rocket plus the weight of the cab:

$$G = Q_0 + g_K,$$

where Q_0 - weight, isolated into all m of step/stages.

Each step/stage has some fuel reserve. After fuel depletion, the waste step/stage is discarded and enters into action following.

The velocity of rocket at the end of the active section W is composed of m of velocity increments $w_1, w_2, \ldots, w_m$, which it it acquires on the individual sections of trajectory, as a result of the work of each step/stage. The additional velocity w_i, given to rocket at the i step/pitch, depends, in the first place, on the weight X_i, isolated into the i-th step/stage, and in the second place, on that passive weight P, which is necessary to carry this step/stage:

$$w_i = f(X_i, P). \qquad (9.6)$$

It is required to find this weight distribution Q_0 according to separate step/stages, by which the velocity at the end of the active section is maximum.

Solution. Let us consider m of the step/stages of rocket as m of the stages of acceleration. State S of the system prior to the beginning of each step/pitch we will characterize one parameter Q - by remaining weight, which are subject to distribution between step/stages. Control on the i step/pitch consists of the selection of weight X_i, abstract/removed from the remaining weight Q to this, i-th step/stage.

Since a velocity increment, according to formula (9.6), depends on two arguments - weight of step/stage and passive weight P, let us determine this passive weight. It is obvious, it is equal to $P=Q-X_i+g_K$, and a velocity increment will be:

$$w_i=f(X_i,\ Q-X_i+g_K).$$

Under the effect of control X_i the system passes from state Q into state $Q'=Q-X_i$.

fundamental functional equation will take the form:

$$W_i(Q)=\max_{0<X_i<Q}\{f(X_i,\ Q-X_i+g_K)+W_{i+1}(Q-X_i)\}. \qquad (9.7)$$

Optimum control at the i step/pitch is the value $X_i=x_i$, with which it is reached this maximum.

Optimum control at the m step/pitch (under the natural

assumption that with a gain in weight, abstract/removed by step/stage, a velocity increment it increases), lies in the fact that, weighing out to last/latter step/stage of entire remaining weight Q. In this case, at last/latter step/pitch, there will be acquired the velocity:

$$W_m(Q) = f(Q, g_K). \qquad (9.8)$$

Further the procedure of dynamic programming is run up/turned by usual order. As a result is located the optimum set of the weights of the step/stages:

$$x = (x_1, x_2, \ldots, x_m), \qquad (9.9)$$

imparting to last/latter step/stage (cab) the maximum speed:

$$W_{max} = W_1(Q_0).$$

Page 175.

10. Problems of dynamic programming ~~s ly~~ with multiplicative criterion.

Until now, we examine only such problems of the dynamic programming in which the gain (criterion, or the index of efficiency) is composed of the sum of gains w_i at the separate step/pitches:

$$W = \sum_{i=1}^{m} w_i, \qquad (10.1)$$

i.e. was additive.

Sometimes appear the problems, in which value W represents by itself not sum, but the product:

$$W = \prod_{i=1}^{m} w_i, \qquad (10.2)$$

AD-A066 460 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/2
OPERATIONS RESEARCH. PART I,(U)
JUN 78 Y S VENTTSEL'
UNCLASSIFIED FTD-ID(RS)T-0687-78-PT-1 NL

5 OF 6
AD
A066460

where w_i - a gain at the i step/pitch (it is assumed that everything w_i are positive). This index or the criterion of efficiency is called multiplicative.

It is not difficult to ascertain that any problem with multiplicative criterion can be reduced to problem with additive criterion. For this, is sufficient, for example, to take the logarithm of expression (10.2) and to seek the solution, which rotates into maximum the logarithm of value W. Since logarithm - increasing function, then the maximum of logarithm corresponds to the maximum of value W.

However, for the solution of problems with multiplicative criterion, there is no direct/straight necessity to without fail take the logarithm of it. Entire/all procedure of dynamic programming can be for this case constructed directly. As the basis its is put this selection of conditional optimum control at each step/pitch by which is converted into maximum the gain at all remaining step/pitches, equal to the product of gain at this step/pitch and of the already optimized gain at all subsequent step/pitches.

The fundamental functional equation of dynamic programming for

this case will take the form:

$$W_i(S) = \max_{U_i} \{w_i(S, U_i) \cdot W_{i+1}(\varphi_i(S, U_i))\}. \quad (10.3)$$

and the condition of the optimum character of last/latter step/pitch will be preserved in the same form as with the additive criterion:

$$W_m(S) = \max_{U_m} \{w_m(S, U_m)\}. \quad (10.4)$$

Entire/all procedure of dynamic programming with multiplicative criterion in no way differs from the usual, besides the fact that under the sign of maximum stands not the sum, but product.

Let us consider one of the typical problems of dynamic programming with multiplicative criterion.

Page 176.

Distribution of means for the increase of the reliability of technical equipment/device.

There is the technical equipment/device S, which consists of m of cell/elements, or node/units $Э_1, Э_2, \ldots Э_m$ (see Fig. 3.39). The failure-free operation of each cell/element is unconditionally necessary for the work of equipment/device S as a whole.

Cell/elements can reject (go out of order), moreover independently of each other. The reliability (probability of

failure-free operation) of entire equipment/device is equal to the product of the reliability of all cell/elements:

$$P = \prod_{i=1}^{m} p_i, \qquad (10.5)$$

where p_i - reliability of the i cell/element.

Available are some means K_0 (in, weight or other monetary terms), which can be used to the increase of the reliability of cell/elements.

A quantity of means X, inserted into the attachments, which increase the reliability of the i cell/element, leads it to the value

$$p_i = f_i(X). \qquad (10.6)$$

All the functions $f_i(X)$ - nondecreasing.

It is required to determine the optimum distribution of means according to cell/elements, which leads to the greatest reliability of equipment/device as a whole.

Problem is solved by the method of dynamic programming. Before us - problem with multiplicative criterion. gain at the i step/pitch $p_i = f_i(X_i)$, where the control X_i - quantity of means, inserted into the i cell/element.

Fundamental functional equation takes the form:

$$P_i(K)=\max_{0<X_i<K}\{f_i(X_i)\cdot P_{i+1}(K-X_i)\}, \quad (10.7)$$

where $P_i(K)$ - a conditional optimum gain, i.e., the maximum reliability of the equipment/device, comprised of all cell/elements, beginning with the i-th and to the m-th, if after i - the 1st step/pitch, i.e., after the provision with the means of previous i - 1 cell/elements, available remained means K. Conditional optimum control at the i step/pitch $x_i(K)$ - the quantity of means at which is reached this maximum.

As in all problems of distributing the rescurce/lifetimes where the means are expend/consumed to end, and gain - nondecreasing function, optimum control on last/latter step/pitch lies in the fact that, isolating into this step/pitch of all the remaining means:

$$x_m(K)=K. \quad (10.8)$$

In this case, is reached the conditional cptimum gain, equal to

$$P_m(K)=f_m(K). \quad (10.9)$$

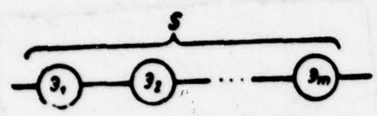


Fig. 3.39.

Page 177.

By the consecutive application/use of formula (10.7) for i = m - 1, m - 2, ..., 2, as ever, we find the conditional optimum controls

$$x_{m-1}(K),\quad x_{m-2}(K),\ \ldots,\ x_2(K)$$

and conditional optimum gains

$$P_{m-1}(K),\quad P_{m-2}(K),\ \ldots,\ P_2(K).$$

The first step/pitch in this case is optimized not conditionally, but it is unconditional, since an initial quantity of means K_0 is assigned:

$$P_1(K_0)=\max_{0\leqslant X_1\leqslant K_0}\{f_1(X_1)\cdot P_2(K_0-X_1)\}. \qquad (10.10)$$

The control

$$x_1=x_1(K_0),$$

at which is reached maximum (10.10), and there is unconditional optimum control at the first step/pitch, but $P_1(K_0)$ - unconditional optimum gain, i.e., maximally attainable by given means the reliability of equipment/device. Further optimum control is constructed according to the diagram:

$$x_1\to K_1^*=K_0-x_1\to x_2\to K_2^*=K_1^*-x_2\to\ldots$$
$$\ldots\to K_{m-1}^*=K_{m-2}^*-x_{m-1}\to x_m\to K_m^*=0.$$

11. Infinite-step process of dynamic programming.

All problems of the dynamic programming which we examine, until now, are related to the processes, which were being divided into the finite Mach number of step/pitches. It goes without saying that all the practical problems, connected with gliding/planning of economic and similar to them operations, are related to this class - to plan/glide it makes sense only to the finite (even very large) segment of time forward. However, there are the problems, in which this section of time is represented previously not by completely determined, and us it can interest the solution of the problem of optimum gliding/planning irrespectively of that, at which precisely step/pitch the operation will be finished. In such cases sometimes there is expedient to consider as the model of phenomenon certain idealized infinite-step controlled process, which will be obtained from real with $m \to \infty$. This model is convenient in that in it there is of exceptional on its role "last/latter step/pitch" - all the step/pitches between themselves are equal, no process in known sense uniform. Conditional optimum control in this process proves to be not depending on the number of step/pitch, but depending only on state S of system S prior to the beginning of step/pitch. It goes without saying that for this it is necessary that the step/pitches would be

uniform, i.e., the functions, determining income and change in the state of system under control effect, were for all step/pitches identical.

Page 178.

One should emphasize that in uniform infinite-step process identical for all step/pitches prove to be only conditional optimum controls; as concerns unconditional optimum control, then it, being dependent on the state of system, reached to this step/pitch, in the general case varies from one step/pitch to the next.

Let us note that unlike finite-step problems, for which optimum control always exists, infinte-step problems can and not have solution. In order to be convinced of this, let us consider an elementary example.

Let there be the problem of distributing the resource/lifetimes with the redundancy (see § 6), but with the infinite number of step/pitches. Means X, inserted into production, give for year income f(X) and are expend/consumed to end. Available is an initial supply of means K_0, which is required optimally to distribute on years, so as to total income would be maximum.

Existence and the form of solution depends on which form of the function f (X).

Let us assume that this function is convex downward (Fig. 3.40). Then it is obvious that the optimum solution exists and lies in the fact that, putting into production of all the available means in the first year. It is real/actual, let us suppose that we, for example, divided means in half, the first half they put in production on the first year, and the second - on following year. It is obvious, this will be disadvantageous, since for the convex downward function f (X)

$$2f(K_0/2) < f(K_0).$$

Let us assume now that function f (X) is convex upward (Fig. 3.41).

it is obvious that in this case it is profitable not to pack into production all means immediately, but "to stretch" them. For example, if we, instead of packing into production of all means at the first step/pitch, is distributed them to two step/pitch, then we will obtain the larger income:

$$2f(K_0/2) > f(K_0),$$

to three step/pitch - still larger, and so forth. With an increase in the number of step/pitches, to which are distributed the means, income only grows.

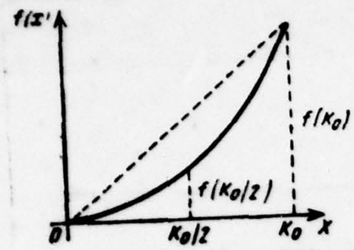


Fig. 3.40.

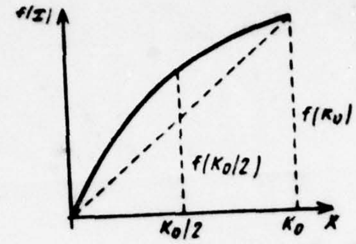


Fig. 3.41.

Page 179.

Let us determine limit, to which will strive total income with the unlimited increase of the number of step/pitches, into which are packed the means, and that means that during the simultaneous decrease of the number of means, packed at each step/pitch. Let us assume first that we plan/glide to m of years and each year we pack into production one and the same sum

$$\Delta X = K_0/m,$$

and then let us direct m to infinity, but ΔX - to zero. Figure 3.42 shows that with sufficiently small ΔX it is possible to replace section curved f(X) with the section of tangent in the beginning of coordinates. Then the income, obtained for year, is approximately equal to

$$f'(0)\,\Delta X \approx f'(0)\,K_0/m,$$

where f'(0) - a value of the derivative of income in the beginning of coordinates. In this case, total income during entire period of m of years will be approximately equal to

$$W \approx f'(0) K_0. \qquad (11.1)$$

With m $\longrightarrow$ - approximate equality (11.1) is converted into precise.

Thus, mu obtained the paradoxical conclusicn/derivation: than lesser means we pack into production on each year, the fact will more be income; within limit, with m $\longrightarrow$ ∞, will be obtained maximum income (11.1). But if we directly pass to limiting case and to place ΔX = 0, i.e., not to pack into the production of any means, then, obviously, and income will be equal to zero.

This is an example of the infinte-step problem where the optimum solution does not exist. With any final m it exists and lies in the fact that, packing of means in all stages equally, while with the infinite number of step/pitches, it ceases to exist.

With setting and solution of infinite-step problems by the method of dynamic programming it is always necessary to trace a question concerning the existence of solution [1].

FOOTNOTE [1]. The conditions for existence of solution in infinite-step

problems are examined, for example, in [10]. ENDFOOTNOTE.

Infinite-step model in the problems of dynamic programming in a series of the cases can render/show simpler than finite-step. It is real/actual, instead of a series of functional equations, solved one after another in the usual procedure of dynamic programming, here it is necessary to solve in all only one the functional equation for a conditional optimum gain, suitable for any step/pitch.

Let us register this only functional equation.

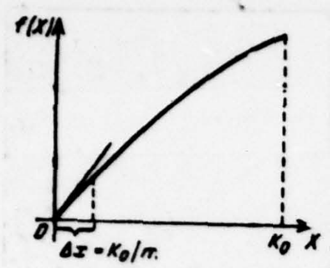


Fig. 3.42.

Page 180.

Let the infinite-step controlled process occur in the physical system $\mathcal{S}$; let us designate S - state of this system after some (any) step/pitch. Under the effect of control U, system $\mathcal{S}$ for the next step/pitch passes into the new state S', which depends on past state S and the used control U:

$$S' = \varphi(S, U).$$

For this step/pitch we obtain gain (income) w, also depending on $\mathcal{S}$ and U:

$$w = f(S, U).$$

Then it is possible to write basic functional equation for infinite-step problem in the form:

$$W(S) = \max_U \{f(S, U) + W(\varphi(S, U))\}, \tag{11.2}$$

where W(S) - the conditional maximum gain which can be obtained, managing system, which is found in state S. In equation (11.2) W(S) - only unknown function; remaining functions (φ, f) are given ones.

Conditional optimum control $u(S)$ - the control at which is reached maximum (11.2).

In some simplest problems succeeds in selecting function $W(S)$ so that it would satisfy equation (11.2). The general methods of the analytical solution of functional equations do not exist. In cases when it is impossible to select function $W(S)$, that satisfies equation (11.2), they resort to approximate solution of this equation. For this, can be used the method of successive approximations, which consists of following: is solved the problem of dynamic programming for final, continually of the increasing number of step/pitches m; if solution exists, then with increase m of function $W_i(S)$ and $u_i(S)$, that determine conditional optimum gain and conditional optimum control for the step/pitches, are sufficient distant from end, they are stabilized, approaching appropriate by functions $W(S)$ and $u(S)$ for infinite-step process, as which they can be approximately undertaken.

In conclusion let us note that the infinite-step problems of dynamic programming can be obtained not only because of the unlimited increase in the number of step/pitches at the assigned length of each step/pitch, but also because of the unlimited decrease of the length of step/pitch Δt, when discrete step by step control passes into continuous. Such problems are fairly complicated, and we will not be on them stopped.

Page 181.

4. SIMULATION OF OPERATIONS ACCORDING TO THE PATTERN OF MARKOVIAN PROCESSES.

1. The Markovian process S by discrete states.

Many operations, which it is necessary to analyze at the visual angle of the selection of optimum solution, are developed as random processes, course and issue of which depend on a series of the random factors, accompanying these operations.

In order to compute the numerical parameters, which characterize the efficiency of such operations, it is necessary to construct certain probabilistic model of phenomenon, which considers its accompanying random factors.

For the mathematical description of many operations, which develop in the form of random process, can be successfully used the mathematical apparatus, worked out in the probability theory for the so-called Markovian processes.

Let us explain the concept of the Markovian process.

Let there be certain physical system S whose state varies in the course of time (under system S can be understood anything: technical equipment/device, repair shop, computer, railroad junction, etc.). If the state of system S varies in time by random, previously unpredicted/unpredictable form, we say that in system S proceeds the random process.

Examples of random processes they can be:

- process of functioning of ETsVM [ЭЦВМ - digital computer];

- process of guidance to the target/purpose of the guided missile or space vehicle;

- process of maintain/serviceing the clients of barbershop or repair shop;

- process of the fulfilment of the plan of the supply of the group of enterprises, etc.

The concrete/specific/actual course of each of such processes depends on a series of the random, previously unpredicted/unpredictable factors, such as:

- admission of orders by ETsVM and the form of these orders; random output/yields of ETsVM from system;

- the random disturbances (interference) in the system of rocket control;

- random character of the flow of the claims (requirements), of the entering from the side clients;

- random interruptions in the fulfilment of the plan of supply, etc.

Page 182.

The random process, which takes place in system S, is called Markov process (or "process without aftereffect"), if it possesses the following property:

For each torque/moment of time t_0 probability of any state of system in the future (with $t > t_0$) depends only on its state in

present (with t = t_0) and dces not depend cn that, when and how system arrived into this state (i.e. as it was developed process in the past).

In other words, in the Markovian process its future development depends only on present state and does not depend on the "prehistory" cf process.

Let us consider an example. Let system **S** represent technical equipment/device, which already studied certain time, by corresponding form "was worn out" and arrived into certain state, which was being characterized by the certain degree of worn out nature S. Us it interests, as will work system in the future. It is clear that, at least in the first apprcximation, the performance characteristics of system in the future (failure rate, the necessity for repair) depend on the state of equipment/device at present torque/moment and do not depend on that, when and as equipment/device it achieved its present state.

In practice frequently are encountered the random processes, which, with one or the other degree of approximation, can be considered Markov.

The theory of the Markovian processes is the at present very

vast section of the probability theory with the wide spectrum of different application/appendices - from the description of physical phenomena of the type of diffusion or mixing of charge during smelting in blast furnace to the processes of queueing or propagation of the mutations of genes in biological population. Us will interest, mainly, the application/uses of theory of the Markovian processes to the construction of the mathematical models of the operations, course and issue of which depends substantially on random factors.

The Markovian processes are divided into classes according to some sign/criteria, depending on [redacted] and at which moment of time system **S** can vary its states.

Random process is called process with discrete states, if the possible states of the system:

$$S_1, S_2, S_3, \ldots$$

can be enumerated (to index one after another, and process itself lies in the fact that from time to time system **S** abruptly (instantly) jumps of one state into another.

Page 183.

Example 1. The technical equipment/device **S** consists of two assemblies: I and II, each of which can during the work of

equipment/device fail (break down). Are possible the following states of the system:

S_1 - both assembly work;

S_2 - first assembly failed, the second works;

S_3 - second assembly failed, the first works;

S_4 - both assemblies failed.

The process, which takes place of system, lies in the fact that it randomly, at some moment of time, passes (it jumps) from state into state. In all the system has four possible states which we will index. Before us - process with discrete states.

Besides processes with discrete states, there are random processes with the continuous states: for these processes is characteristic gradual, smooth transition from state into state. For example, the process of changing the voltage in lighting system represents by itself random process with continuous states.

In this chapter we will examine only random processes with discrete states.

During the analysis of random processes with discrete states, it is very convenient to use geometric pattern - by the so-called graph/count of states. The graph/count of states geometrically represents the possible states of system and its allowed transitions from state into state.

Let there be system S with the discrete states:

$$S_1, S_2, \ldots, S_n.$$

We will represent each state as rectangle, and allowed transitions ("jump/migrations") from state into state - by rifleman/pointers, who combine these rectangles (Fig. 4.1).

Let us note that by rifleman/pointers are noted only direct transitions from state into state; if system can pass from state S_1 in S_3 only through S_2, then by rifleman/pointers will be noted only transitions $S_1 \longrightarrow S_2$ and $S_2 \longrightarrow S_3$< but not $S_1 \longrightarrow S_3$.

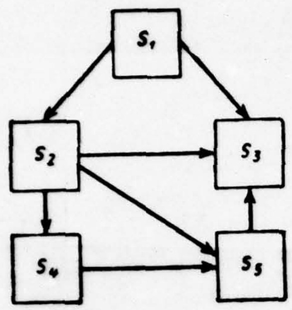


Fig. 4.1.

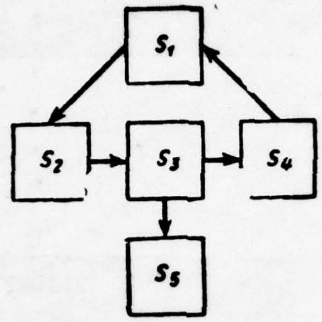


Fig. 4.2.

Page 184.

Example 2. S - truck system which can be located in one of the five possible states:

S_1-is exact, it works;

S_2-is defective, it expects inspection;

S_3 - will be scanned;

S_4 - is overhauled;

S_5 is copied.

The graph/count of the states of system is shown on Fig. 4.2.

Example 3. To construct the graph/count of states under conditions of example 1 (is assumed that the repair of assemblies in the course of process is not produced).

Solution. The graph/count of states is represented in Fig. 4.3. Let us note that on graph/count is not shown the allowed transition from state S_1 directly in S_4 which will be carried out, if strictly simultaneously leave the system both assembly. The possibility of this event we disregard.

Example 4. System S, as in example 1, represents technical equipment/device, which consists of two assemblies: I and II; each of them can at some moment of time refuse. The refused assembly immediately begins to be restored.

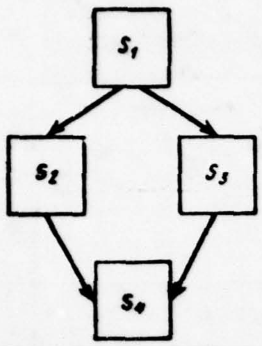


Fig. 4.3.

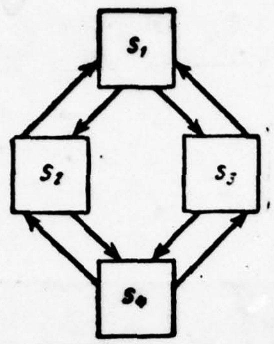


Fig. 4.4

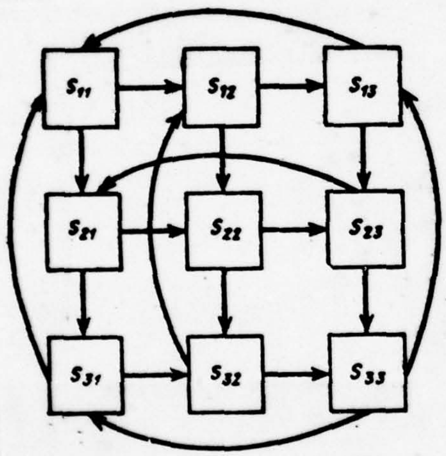


Fig. 4.5.

Page 185.

The possible states of the system:

S_1 - both assembly work;

S_2 - first assembly is restored, the second works;

S_3 - first assembly works, the second is restored;

S_4 - both assembly are restored.

The graph/count of the states of system is shown on Fig. [redacted] 4.4.

Example 5. Under conditions of example [redacted] 4 each assembly before beginning to be restored, undergoes inspection for purpose of localization of malfunction.

The states of system let us for convenience label not by one, but by two indices; the first will indicate the states of the first assembly:

1 - works,

2 - will be scanned,

3 - are restored;

the second - the same states for the second assembly, so that, for example, S_{23} will indicate; the first assembly it will be scanned, the second - is restored, and so forth.

The possible states of system **S** will be:

S_{11} - both unit work,

S_{12} - first unit works, the second will be scanned,

. . . .

S_{33} - both unit are restored.

(a total of 9 states).

The graph/count of states is shown on Fig. 4.5.

2. Random processes ~~S by~~ with discrete and by continuous time. Markov target/purpose.

The methods of the mathematical description of the Markovian process, which takes place in system with discrete states, depend on that, at which torque/moments of time - previously known or random -

can occur the transitions ("jump/migrations") of system from state into state.

random process is called process with discrete time, if the transitions of system from state into state are possible only at the strictly defined, previously fixed/recorded moment of the time: t_1, t_2, ... In time intervals between these torque/moments system S retains its state.

R
random process is called process with continuous time, if the transition of system from state into state is feasible in the any, in advance unknown, random moment t.

Let us consider first of all the Markovian process with discrete states and discrete time.

Let there be the physical system S, which can be located in the states:

$$S_1, S_2, \ldots, S_n,$$

moreover the transitions ("jump/migrations") of system from state into state cart are possible only at the torque/moments:

$$t_1, t_2, \ldots, t_k, \ldots.$$

Let us call these torque/moments the "step/pitches" or by the "stages" of process and examine the random process, which occurs in

system S as function of the integral argument: 1, 2, ..., k, ... (number of step/pitch).

Page 186.

The random process, which occurs of system, lies in the fact that at successive moments of time t_1, t_2, t_3, ..., system S proves to be in one or the other states, behaving, for example, as follows:

$$S_1 \to S_3 \to S_5 \to S_4 \to S_2 \to S_1 \to \ldots$$

or

$$S_1 \to S_2 \to S_1 \to S_2 \to S_3 \to S_4 \to S_1 \to \ldots.$$

In general case at torque/moments t_1, t_2, ... system can not only vary state, but also remain in previous, for example:

$$S_1 \to S_1 \to S_2 \to S_3 \to S_3 \to S_3 \to S_4 \to S_1 \to \ldots$$

Let us agree to designate $S_i^{(k)}$ the event, which consists of the fact that after k of step/pitches the system is in state S_i. With any k of the event

$$S_1^{(k)}, S_2^{(k)}, \ldots, S_i^{(k)}, \ldots, S_n^{(k)}$$

is formed full/total/complete group and are incompatible/inconsistent.

The process, which occurs in system, can be presented as sequence (chain/network) of the events, for example:

$$S_1^{(0)}, S_2^{(1)}, S_1^{(2)}, S_2^{(3)}, S_3^{(4)}, \ldots$$

This random sequence of events is called Markov chain, if for each step/pitch transitional probability from any state S_i into any S_j does not depend on that, when and as system it arrived into state S_i.

We will describe Markov chain with the help of the so-called probabilities of states. Let at the any moment of time (after any, k step/pitch) system S can be in one of the states:

$$S_1, S_2, \ldots, S_n,$$

i.e. it will be carried out one of the full/total/complete group of the antithetical events:

$$S_1^{(k)}, S_2^{(k)}, \ldots, S_n^{(k)}.$$

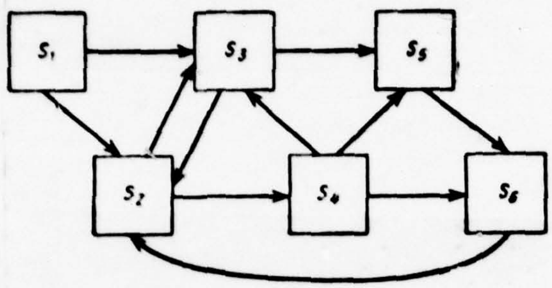


Fig. 4.6.

Page 187.

Let us designate the probabilities of these events:

$$p_1(1) = P(S_1^{(1)}); \quad p_2(1) = P(S_2^{(1)}); \ \ldots; \ p_n(1) = P(S_n^{(1)})$$

- probability after the first step/pitch.

$$p_1(2) = P(S_1^{(2)}); \quad p_2(2) = P(S_2^{(2)}); \ \ldots; \ p_n(2) = P(S_n^{(2)}) \quad (2.1)$$

- probability after the second step/pitch; and generally after the k step/pitch:

$$p_1(k) = P(S_1^{(k)}); \quad p_2(k) = P(S_2^{(k)}); \ \ldots; \ p_n(k) = P(S_n^{(k)}). \quad (2.2)$$

It is easy to see that for each number of step/pitch k

$$p_1(k) + p_2(k) + \ \ldots \ + p_n(k) = 1,$$

since this - the probability of incompatible events, which form full/total/complete group.

Let us call the probabilities

$$p_1(k), \ p_2(k), \ \ldots, \ p_n(k)$$

the probabilities of states; let us assign the mission: to find the probabilities of the states of system for any k.

It is represented the state of system in the form of graph/count (Fig. 4.6), where the pointers showed possible transitions of system from state into state for one step/pitch.

Random process (Markov chain) can be visualized in the manner that as if the point, which represents system S, randomly moves (it roams) on the graph of states, jumping from state into state at torque/moments t_1, t_2, ..., and sometimes (in the general case) and being detained some number of step/pitches in one and the same state. For example, the sequence of the transitions

$$S_1 \rightarrow S_2 \rightarrow S_2 \rightarrow S_3 \rightarrow S_3 \rightarrow S_4 \rightarrow S_6 \rightarrow S_3$$

it is possible to represent on the graph/count of states as sequence of different positions of the point (see broken pointers, representing transitions from state into state in Fig. 4.7). The "delay" of system in state S_2 on third stage is depicted by the arrow/pointer, outgoing from state S_2 and to it returning.

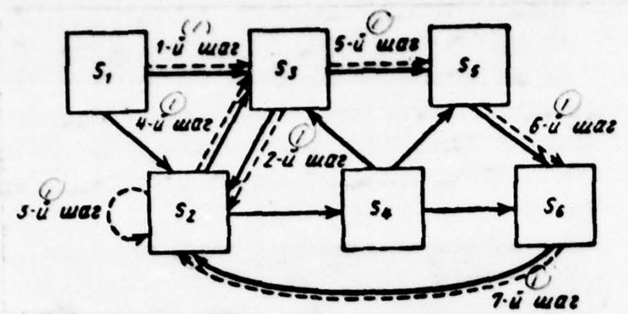


Fig. 4.7.

Key: (1). step/pitches.

Page 188.

For any step/pitch (torque/moment of time $t_1, t_2, \ldots, t_k, \ldots$ or number 1, 2, ..., k, ...) exist some the transitional probabilities of system from any state into any another (some of them are equal to zero, if direct transition for one step/pitch is impossible), and also the probability of the delay of system in this state.

Let us call these probabilities the transient probabilities Markov chain.

Markov chain is called uniform, if transient probabilities do not depend on the number of step/pitch. Otherwise Markov chain is called heterogeneous.

Let us consider first uniform Markov chain. Let system S have n of possible states $S_1, S_2, \ldots, S_n$. Let us assume that for each state to us is known transitional probability into any other state for one step/pitch (including the probability of delay in this state). Let us designate P_{ij} transitional probability for one step/pitch of state S_i of state S_j; P_{ii} it will be the probability of the delay of system in state S_i. Let us register transient probabilities P_{ij} in the form of rectangular array (matrix/die):

$$\|P_{ij}\| = \left\| \begin{array}{cccccc} P_{11} & P_{12} & \ldots & P_{1j} & \ldots & P_{1n} \\ P_{21} & P_{22} & \ldots & P_{2j} & \ldots & P_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{i1} & P_{i2} & \ldots & P_{ij} & \ldots & P_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{n1} & P_{n2} & \ldots & P_{nj} & \ldots & P_{nn} \end{array} \right\|. \qquad (2.3)$$

Some of the transient probabilities P_{ij} can be equal to zero: this means that for one step/pitch the transition of system of the i state into the j-th is impossible. Along the principal diagonal of the matrix/die of transient probabilities, stand the probabilities P_{ii} of the fact that the system will not leave the state S_i, but it will remain in it.

Using introduced above events $S_1^{(k)}, S_2^{(k)}, \ldots, S_n^{(k)}$, transient probabilities P_{ij} can be registered as conditional probabilities:

$$P_{ij} = P(S_j^{(k)}/S_i^{(k-1)}).$$

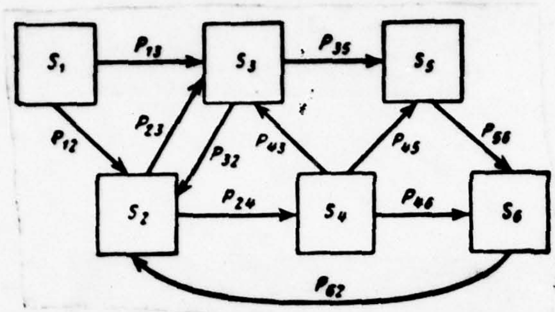


Fig. 4.8.

Page 189.

Hence it follows that the sum of the terms, which stand of each of buildings of matrix/die (2.3), must be equal to unity, since, in whatever state the system was before the k step/pitch, events $S_1^{(k)}, S_2^{(k)}, \ldots, S_n^{(k)}$ were incompatible/inconsistent and form full/total/complete group.

In the examination Markov chains, frequently it is to convenient use the graph/count of states, on whom of arrow/pointers are written the corresponding transition probabilities (see Fig. 4.8). This graph/count we will call the "labeled graph/count of states".

Let us note that in Fig. 4.8 are written not all transient probabilities, but only those of them, that are not equal to zero and vary the state of system, i.e., P_{ij} with i $\neq$ j; "probability of

delay" P_{11}, P_{22}, ... to enter/write on graph/count is excessive, since each of them supplements to unity the sum of the transient probabilities, which correspond to all arrow/pointers, which proceed from this state. For example, for the graph of Fig. 4.8

$$P_{11}=1-(P_{12}+P_{13}), \quad P_{44}=1-(P_{43}+P_{45}+P_{46}),$$
$$P_{22}=1-(P_{23}+P_{24}), \quad P_{55}=1-P_{56},$$
$$P_{33}=1-(P_{32}+P_{34}), \quad P_{66}=1-P_{63}.$$

If from state S_i proceeds not one arrow/pointers (transition from it not into which another state it is impossible), the corresponding probability of delay P_{ii} it is equal to unity.

Having available the labeled graph/count of states (or, that equivalently, the matrix/die of transient probabilities) and knowing the initial state of system, it is possible to find the probabilities of the states

$$p_1(k),\ p_2(k),\ \dots,\ p_n(k)$$

after any (k-th) step/pitch.

Let us show how that is made.

Let us assume that at the initial moment (before the first step/pitch) the system is in some specific state, for example, Then, for initial moment (0) we are have:

$$p_1(0)=0; \quad p_2(0)=0; \dots; \quad p_m(0)=1; \dots; \quad p_n(0)=0,$$

i.e. the probability of all states are equal to zero, besides the

probability of initial state S_m, which is equal to unity.

Let us find the probabilities of states after the first step/pitch. We know that before the first step/pitch the system knowingly is in state S_m. That means that for the first step/pitch it will pass into states $S_1, S_2, \ldots, S_m, \ldots, S_n$ with the probabilities

$$P_{m1}, P_{m2}, \ldots, P_{mm}, \ldots, P_{mn},$$

those registered in the m matrix row of transient probabilities. Thus, the probabilities of states after the first step/pitch will be:

$$p_1(1)=P_{m1};\quad p_2(1)=P_{m2}; \ldots;\quad p_m(1)=P_{mm}; \ldots;\quad p_n(1)=P_{mn}. \quad (2.4)$$

Page 190.

Let us find the probabilities of states after the second step/pitch:

$$p_1(2),\quad p_2(2), \ldots,\quad p_i(2), \ldots,\quad p_n(2).$$

Let us compute them according to the formula of composite probability, with the hypotheses:

- after the first step/pitch system was in state S_1;

- after the first step/pitch system was in state S_2;

. . . .

- after the first step/pitch system will be in state S_i;

. . . .

- after the first step/pitch system will be in state S_n.

The probabilities of hypotheses are known (see (2.4)); the conditional probabilities of transition into state S_i with each hypothesis are also known and registered in the matrix/die of transient probabilities. On the formula of composite probability, we will obtain:

$$\left.\begin{array}{l} p_1(2)=p_1(1)P_{11}+p_2(1)P_{21}+\ldots+p_n(1)P_{n1}; \\ p_2(2)=p_1(1)P_{12}+p_2(1)P_{22}+\ldots+p_n(1)P_{n2}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_i(2)=p_1(1)P_{1i}+p_2(1)P_{2i}+\ldots+p_n(1)P_{ni}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_n(2)=p_1(1)P_{1n}+p_2(1)P_{2n}+\ldots+p_n(1)P_{nn}, \end{array}\right\} \quad (2.5)$$

or, it is much shorter,

$$p_i(2)=\sum_{j=1}^{n} p_j(1)P_{ji} \quad (i=1,\ldots,n). \quad (2.6)$$

In formula (2.6) the addition extends formally to all states $S_1, \ldots, S_n$; to actually consider is necessary only those of them, for which transient probabilities P_{ji} are different from zero, i. e., those states from which can be completed the transition into state S_i (or delay in it).

Thus, the probabilities of states after the second step/pitch

are known. It is obvious, after the third step/pitch they are determined analogously:

$$p_i(3)=\sum_{j=1}^{n} p_j(2)\,P_{ji}, \qquad (2.7)$$

and generally after the k step/pitch:

$$p_i(k)=\sum_{j=1}^{n} p_j(k-1)\,P_{ji} \qquad (i=1, \ldots, n). \qquad (2.8)$$

Thus, the probabilities of states $p_i(k)$ after the k step/pitch are determined by recursion formula (2.8) through the probabilities of states from (k - 1) step; those, in turn, - through the probabilities of states after (k - 2) step and, etc.

Page 191.

Example 1. On some target/purpose is conducted shooting four shots at the moment of time t_1, t_2, t_3, t_4.

The possible states of target/purpose (system S):

S_1 - target/purpose is unharmed;

S_2 - target/purpose is insignificantly injured;

S_3 - target/purpose will obtain essential damages;

S_4 - target/purpose is completely struck (it cannot function).

The labeled graph/count of the states of system is shown on Fig. 4.9.

At initial moment target/purpose is located in states S_1 (it is not injured). To determine the probabilities of the states of target/purpose after four shots.

Solution. From the graph of states we have:

$$P_{12}=0{,}4;\quad P_{13}=0{,}2;\quad P_{14}=0{,}1 \quad \text{and} \quad P_{11}=1-(P_{12}+P_{13}+P_{14})=0{,}3.$$

Analogously we find:

$$\begin{array}{llll} P_{21}=0; & P_{22}=0{,}4; & P_{23}=0{,}4; & P_{24}=0{,}2; \\ P_{31}=0; & P_{32}=0; & P_{33}=0{,}3; & P_{34}=0{,}7; \\ P_{41}=0; & P_{42}=0; & P_{43}=0; & P_{44}=1. \end{array}$$

Thus, the matrix/die of transient probabilities takes the form:

$$\|P_{ij}\|=\begin{Vmatrix} 0{,}3 & 0{,}4 & 0{,}2 & 0{,}1 \\ 0 & 0{,}4 & 0{,}4 & 0{,}2 \\ 0 & 0 & 0{,}3 & 0{,}7 \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

Since at the initial moment target/purpose S is located in state S_1, then

$$p_1(0)=1.$$

The probabilities of states after the first step/pitch (shot) are taken from the first matrix row

$$p_1(1)=0{,}3;\quad p_2(1)=0{,}4;\quad p_3(1)=0{,}2;\quad p_4(1)=0{,}1.$$

Probabilities of states after the second step/pitch:

$$p_1(2)=p_1(1)\,P_{11}=0{,}3\cdot 0{,}3=\underline{0{,}09};$$

$$p_2(2)=p_1(1)\,P_{12}+p_2(1)\,P_{22}=0{,}3\cdot 0{,}4+0{,}4\cdot 0{,}4=\underline{0{,}28};$$

$$p_3(2)=p_1(1)\,P_{13}+p_2(1)\,P_{23}+p_3(1)\,P_{33}=0{,}3\cdot 0{,}2+0{,}4\cdot 0{,}4+0{,}2\cdot 0{,}3=\underline{0{,}28};$$

$$p_4(2)=p_1(1)\,P_{14}+p_2(1)\,P_{24}+p_3(1)\,P_{34}+p_4(1)\,P_{44}=$$
$$=0{,}3\cdot 0{,}1+0{,}4\cdot 0{,}2+0{,}2\cdot 0{,}7+0{,}1\cdot 1=\underline{0{,}35}.$$

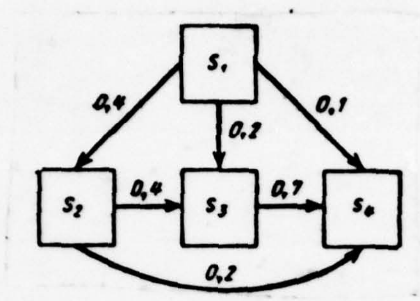


Fig. 4.9.

Page 492.

Probabilities of states after the third step/pitch:

$$p_1(3)=p_1(2)\,P_{11}=0{,}09\cdot 0{,}3=\underline{0{,}027},$$
$$p_2(3)=p_1(2)\,P_{12}+p_2(2)\,P_{22}=0{,}09\cdot 0{,}4+0{,}28\cdot 0{,}4=\underline{0{,}148};$$
$$p_3(3)=p_1(2)\,P_{13}+p_2(2)\,P_{23}+p_3(2)\,P_{33}=$$
$$=0{,}09\cdot 0{,}2+0{,}28\cdot 0{,}4+0{,}28\cdot 0{,}3=\underline{0{,}214};$$
$$p_4(3)=p_1(2)\,P_{14}+p_2(2)\,P_{24}+p_3(2)\,P_{34}+p_4(2)\,P_{44}=$$
$$=0{,}09\cdot 0{,}1+0{,}28\cdot 0{,}2+0{,}28\cdot 0{,}7+0{,}35\cdot 1=\underline{0{,}611}.$$

Probabilities of states after the fourth step/pitch:

$$p_1(4)=p_1(3)\,P_{11}=\underline{0{,}0081};$$
$$p_2(4)=p_1(3)\,P_{12}+p_2(3)\,P_{22}=0{,}27\cdot 0{,}4+0{,}148\cdot 0{,}4=\underline{0{,}0700},$$
$$p_3(4)=p_1(3)\,P_{13}+p_2(3)\,P_{23}+p_3(3)\,P_{33}=$$
$$=0{,}027\cdot 0{,}2+0{,}148\cdot 0{,}4+0{,}214\cdot 0{,}3=\underline{0{,}1288};$$
$$p_4(4)=p_1(3)\,P_{14}+p_2(3)\,P_{24}+p_3(3)\,P_{34}+p_4(3)\,P_{44}=$$
$$=0{,}027\cdot 0{,}1+0{,}148\cdot 0{,}2+0{,}214\cdot 0{,}7+0{,}611\cdot 1=\underline{0{,}7931}.$$

Thus, you obtained the probabilities of all issues of the bombardment of target/purpose (four shots):

- target/purpose is not injured: $p_1(4)\approx 0.008$;

- target/purpose will cbtain the insignificant damages: $p_2(4) \approx 0.070$;

- target/purpose will cbtain the essential damages: $p_3(4) \approx 0.129$;

- target/purpose was struck completely: $p_4(4) \approx 0.793$.

We considered the uniform Markov chain, for which transitional probabilities from one step/pitch to the next do not vary.

Let us consider now the general case - heterogeneous Markov chain for which the probabilities of transition P_{ij} vary from one step/pitch to the next. Let us designate $P_{ij}^{(k)}$ - the transitional probability of system from state S_i into state S_j at the k step/pitch, i. e., the conditional probability

$$P_{ij}^{(k)} = P\left(S_j^{(k)} / S_i^{(k-1)}\right).$$

Let us assume that to us are assigned the matrix/dies of transitional probability at each step/pitch. Then the probability of the fact that system S after k of the step/pitches will be located in state S_i, it will be expressed by the formula:

$$p_i(k) = \sum_j p_j(k-1) P_{ji}^{(k)} \quad (i = 1, \ldots, n), \tag{2.9}$$

which it differs from analogous formula (2.8) for the uniform Markov chain only by the fact that in it they figure transitional probabilities, which depend on the number of step/pitch k. Calculations on formula (2.9) not a bit are not more complex than in the case of uniform circuit.

Page 193.

Example [redacted] 2. Are produced three shots on the target/purpose which can be into the same four states S_1, S_2, S_3, S_4, as in previous example, but transitional probability for three consecutive shots are different and assigned by three matrix/dies:

$$\|P_{ij}^{(1)}\| = \begin{Vmatrix} 0.3 & 0.4 & 0.2 & 0.1 \\ 0 & 0.4 & 0.4 & 0.2 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0 & 1 \end{Vmatrix};$$

$$\|P_{ij}^{(2)}\| = \begin{Vmatrix} 0.1 & 0.4 & 0.3 & 0.2 \\ 0 & 0.2 & 0.5 & 0.3 \\ 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 1 \end{Vmatrix};$$

$$\|P_{ij}^{(3)}\| = \begin{Vmatrix} 0.05 & 0.3 & 0.4 & 0.25 \\ 0 & 0.1 & 0.6 & 0.3 \\ 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

At initial moment target/purpose is located in state S_1. To find the probabilities of states after three shots.

Solution. We have:

$$p_1(1)=0{,}3;\quad p_2(1)=0{,}4;\quad p_3(1)=0{,}2;\quad p_4(1)=0{,}1;$$

$$p_1(2)=p_1(1)\,P_{11}^{(2)}=0{,}3\cdot 0{,}1=\underline{0{,}03};$$

$$p_2(2)=p_1(1)\,P_{12}^{(2)}+p_2(1)\,P_{22}^{(2)}=0{,}3\cdot 0{,}4+0{,}4\cdot 0{,}2=\underline{0{,}20};$$

$$p_3(2)=p_1(1)\,P_{13}^{(2)}+p_2(1)\,P_{23}^{(2)}+p_3(1)\,P_{33}^{(2)}=$$
$$=0{,}3\cdot 0{,}3+0{,}4\cdot 0{,}5+0{,}2\cdot 0{,}2=\underline{0{,}33};$$

$$p_4(2)=p_1(1)\,P_{14}^{(2)}+p_2(1)\,P_{24}^{(2)}+p_3(1)P_{34}^{(2)}\,p_4+(1)\,P_{44}^{(2)}=0{,}3\cdot 0{,}2+$$
$$+0{,}4\cdot 0{,}3+0{,}2\cdot 0{,}8+0{,}1\cdot 1=\underline{0{,}44};$$

$$p_1(3)=p_1(2)\,P_{11}^{(3)}=0{,}03\cdot 0{,}05\approx\underline{0{,}002};$$

$$p_2(3)=p_1(2)\,P_{12}^{(3)}+p_2(2)\,P_{22}^{(3)}=0{,}03\cdot 0{,}3+0{,}20\cdot 0{,}1=\underline{0{,}029};$$

$$p_3(3)=p_1(2)\,P_{13}^{(3)}+p_2(2)\,P_{23}^{(3)}+p_3(2)\,P_{33}^{(3)}=$$
$$=0{,}3\cdot 0{,}4+0{,}20\cdot 0{,}6+0{,}33\cdot 0{,}1=\underline{0{,}165};$$

$$p_4(3)=p_1(2)\,P_{14}^{(3)}+p_2(2)\,P_{24}^{(3)}+p_3(2)\,P_{34}^{(3)}+p_4(2)\,P_{44}^{(3)}=$$
$$=0{,}03\cdot 0{,}25+0{,}20\cdot 0{,}3+0{,}33\cdot 0{,}9+0{,}44.1\approx\underline{0{,}804}.$$

Thus, the probability cf states after three shots:

$$p_1(3)\approx 0{,}002;\quad p_2(3)=0{,}029;\quad p_3(3)=0{,}165;\quad p_4(3)\approx 0{,}804.$$

Page 194.

3. Markov process with discrete states and continuous time. Equations of Kolmogorov for the probabilities of states.

In the previous paragraph we examined Markov chain, i.e., the random process, taking place in the system which randomly can pass from state into state only at some previously specific, fixed/recorded moment of time.

In practice considerably more frequently are encountered the situations when the transitions of system from state into state occur not into those fix/recorded, but at the random moment of time which to previously indicate is impossible - transition it can be carried out, generally speaking, at any moment. For example, breakdown (failure) of any cell/element of equipment can occur at the any moment of time; the termination of the repair (restoration/reduction) of this cell/element also can occur into previously the not fixed torque/moment, etc.

For describing such processes in a series of the cases, can be successfully used the diagram of the Markovian process by discrete states and continuous time, which we will for brevity call continuous Markov chain.

Let us show how are expressed the probabilities of states for this process.

Let there be a series of the discrete states:

$$S_1, S_2, \ldots, S_n;$$

the transition (jump/migration) of system S from state into state can be realized at the any moment of time. The graph/count of the states of system is represented in Fig. 4.10.

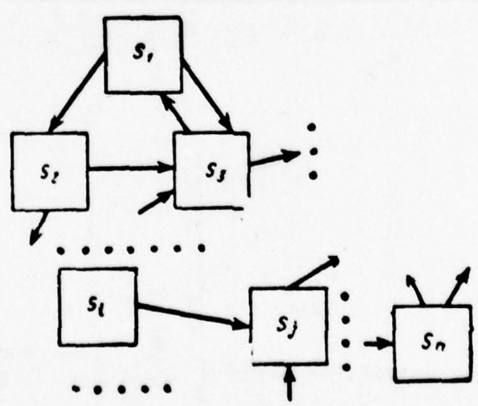


Fig. 4.10.

Page 195.

Let us designate $p_i(t)$ - probability that at torque/moment t system S will be located able S_i (i = 1, ..., n). It is obvious, for any moment t, the sum of the probabilities of states is equal to one:

$$\sum_{i=1}^{n} p_i(t) = 1, \qquad (3.1)$$

since the events, which consist of the fact that at torque/moment t the system is in states S_1, S_2, ..., S_n, are incompatible/inconsistent and form full/total/complete group.

Let us assign mission - to determine for any t of the probability of the states:

$$p_1(t),\ p_2(t),\ \ldots,\ p_n(t).$$

For finding of these probabilities, it is necessary to know the characteristics of process, analogous to transient probabilities for Markov chain. In the case of process with continuous time for us it is not necessary to assign those defined, different from zero, transient probabilities P_{ij}; the probability of transiting (jump/migration) the system from state into state accurately at

torque/moment t will be equal to zero (just as the probability of any separate value of continuous random variable). Instead of transient probabilities P_{ij} we will introduce into the examination of the probability density of transition λ_{ij}.

Let system S at torque/moment t is be in state S_i. Let us consider the elementary time interval Δt, which adjoins torque/moment t (Fig. 4.11).

Let us name the probability density of transition λ_{ij} the limit of relation the transitional probability of system for time Δt from state S_i into state S_j to the length of interval/gap Δt:

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t}, \tag{3.2}$$

where $P_{ij}(\Delta t)$ - probability that the system, which was being located at torque/moment t in state S_i, for time Δt will pass from it into state S_j (probability density of transition it is determined only for j ≠ i).

From formula (3.2) it follows that with small Δt the probability of transition $P_{ij}(\Delta t)$ (with an accuracy to infinitesimal higher orders) is equal to $\lambda_{ij}\Delta t$:

$$P_{ij}(\Delta t) \approx \lambda_{ij}\, \Delta t.$$

If all the probability densities of transition λ_{ij} do not depend

on t (i.e. from that, at which torque/moment it begins elementary section Δt), Markov process is called uniform; if these densities represent by themselves some functions of time $\lambda_{ij}(t)$, process it is called heterogeneous.

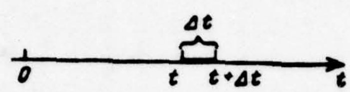


Fig. 4.11.

Page 96.

With the use of the abbreviated name "continuous Markov chain" we also will distinguish uniform and heterogeneous circuits.

Let us assume that to us are known to the probability density of transition λ_{ij} for all pairs of states S_i, S_j.

Let us construct the graph/count of the states of system S and against each arrow/pointer will write the appropriate probability density of transition (Fig. 4.12).

This graph/count, with the written of arrow/pointers probability densities of transition, we will call the labeled graph/count of states.

It proves to be, knowing the labeled graph/count of states, it is possible to define the probabilities of the states:

$$p_1(t),\ p_2(t),\ \ldots,\ p_n(t) \qquad (3.3)$$

as functions of time. Namely, these probabilities satisfy the specific form of differential equations, the so-called equations of Kolmogorov. Solving these equations, we will obtain probabilities (3.3).

Let us demonstrate the methodology of the derivation of the equations of Kolmogorov for the probabilities of states based on specific example.

Let system S have four possible states:

$$S_1,\ S_2,\ S_3,\ S_4;$$

the labeled graph/count of the states of system is shown on Fig. 4.13.

Let us assign to itself the mission: to find one of the probabilities of the states, for example, $p_1(t)$. This be probability that at torque/moment t the system will be located able S_1.

Let us give t a small increase Δt and will find probability that at torque/moment t + Δt the system will be it is located able S_1. How this event it can occur?

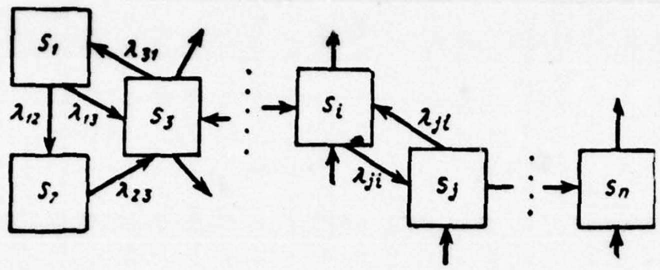


Fig. 4.12.

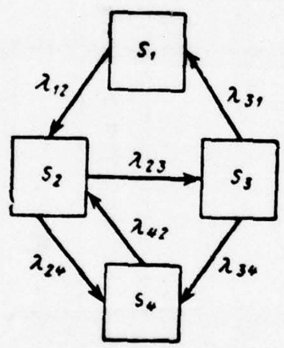


Fig. 4.13.

Page 197.

It is obvious, by two methods:

- at torque/moment t system was already able S_i, but for time

Δt, it did not leave this state or

- at torque/moment t system was with state S_3, but for time Δt, it passed from it into S_1.

The probability of the first version let us find as product of probability $p_1(t)$ that that at torque/moment t the system was able S_1; to conditional probability that, being in state S_1, system for time Δt it will not pass from it into S_2. This conditional probability (with an accuracy to infinitesimal higher orders) is equal to $1 - \lambda_{12}\Delta t$.

It is analogous, the probability of the second version is equal to the probability of the fact that at torque/moment t the system was able S_3, multiplied by the conditional probability of transition for time Δt into state S_1:

$$p_3(t)\,\lambda_{31}\,\Delta t.$$

Employing the rule of addition of probabilities, we obtain:

$$p_1(t+\Delta t)=p_1(t)(1-\lambda_{12}\Delta t)+p_3(t)\lambda_{31}\Delta t.$$

Let us discover brackets in right side, let us transfer $p_1(t)$ into left and let us divide both parts of the equality on Δt; we will obtain:

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t}=-\lambda_{12}p_1(t)+\lambda_{31}p_3(t).$$

Now let us direct Δt to zero and will pass to the limit:

$$\lim_{\Delta t\to 0}\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t}=-\lambda_{12}p_1(t)+\lambda_{31}p_3(t).$$

Left side is nothing else but derivative $p_1(t)$:

$$\frac{dp_1(t)}{dt}=-\lambda_{12}p_1(t)+\lambda_{31}p_3(t). \qquad (3.4)$$

Thus, deduced differential equation which must satisfy function $p_1(t)$. Analogous differential equations can be deduced, also, for the remaining probabilities of the state: $p_2(t)$, $p_3(t)$, $p_4(t)$.

Let us consider the second state S_2 and will find $p_2(t+\Delta t)$ - the probability of the fact that at torque/moment $(t+\Delta t)$ system S will be located able S_2. This event can occur by the following methods:

- at torque/moment t system was already able S_2, but for time Δt, it passed from it either in S_3 or in S_4;

- at torque/moment t system was able S_1, but for time Δt, it passed from it into S_2; or

- at torque/moment t system was able S_4, but for time Δt, it passed from it into S_2.

The probability of the first version is computed as follows: $p_2(t)$ it is multiplied by conditional probability that the system after Δt will pass either in S_3 or in S_4.

Page 198.

Since the events, which consist of transition for time Δt of S_2 of S_3 and of S_2 of S_4, are incompatible/inconsistent, then is probability that to be carried out one of these transitions, is equal to the sum of their probabilities, i.e., $\lambda_{23}\Delta t + \lambda_{24}\Delta t$ (with an accuracy to infinitesimal higher orders). Probability that will be carried out not one their these transitions, is equal to $1 - \lambda_{23}\Delta t - \lambda_{24}\Delta t$. Hence the probability of the first version:

$$p_2(t)(1-\lambda_{23}\Delta t-\lambda_{24}\Delta t).$$

Adjoining here the probability of the second and third versions, we will obtain:

$$p_2(t+\Delta t)=p_2(t)(1-\lambda_{23}\Delta t-\lambda_{24}\Delta t)+ \\ +p_1(t)\lambda_{12}\Delta t+p_4(t)\lambda_{42}\Delta t.$$

Transferring $p_2(t)$ to left side, Dale on Δt and passing to limit, we will obtain differential equation for $p_2(t)$:

$$\frac{dp_2(t)}{dt}=-\lambda_{23}p_2(t)-\lambda_{24}p_2(t)+\lambda_{12}p_1(t)+\lambda_{42}p_4(t). \qquad (3.5)$$

Discussing analogously for states S_3, S_4, we will obtain as a result the system of the differential equations, comprised according

to type (3.4) and (3.5). Let us reject/throw in them for brevity argument t of functions p_1, p_2, p_3, p_4 and will rewrite this system in the form:

$$\left.\begin{aligned}
\frac{dp_1}{dt} &= -\lambda_{12} p_1 + \lambda_{31} p_3, \\
\frac{dp_2}{dt} &= -\lambda_{23} p_2 - \lambda_{24} p_2 + \lambda_{12} p_1 + \lambda_{42} p_4, \\
\frac{dp_3}{dt} &= -\lambda_{31} p_3 - \lambda_{34} p_3 + \lambda_{23} p_2, \\
\frac{dp_4}{dt} &= -\lambda_{42} p_4 + \lambda_{24} p_2 + \lambda_{34} p_3.
\end{aligned}\right\} \quad (3.6)$$

These equations for the probabilities of states are called the equations of Kolmogorov.

Integration of this system of equations will give to us the unknown probabilities states as of function of time. Initial conditions are taken depending on was how the initial state of system S. For example, if at zero time (with t = 0) system S was in state S_1, then it is necessary to take the initial conditions:

$$\text{with} \quad t=0 \quad p_1=1, \quad p_2=p_3=p_4=0.$$

Let us note that all four equations for p_1, p_2, p_3, p_4 it would be possible and not to write; real/actually, $p_1 + p_2 + p_3 + p_4 = 1$ for all t, and any of the probabilities p_1, p_2, p_3, p_4 it is possible

to express by three others. For example p_4 it is possible to express by p_1, p_2, p_3 in the form

$$p_4 = 1 - (p_1 + p_2 + p_3)$$

and to substitute in the remaining equations.

Page 199.

Then special equation for probability p_4 it is possible and not to write. However, subsequently to us it will more conveniently use the full/total/complete system of equations of type (3.6).

Let us focus attention on the structure of equations (3.6). They all are constructed according to the completely specific rule which can be formulated as follows.

On the left side of each equation is the derivative of the probability of state, and right side contains as many members, as arrow/pointers are connected with this state. If arrow/pointer is directed from state, the corresponding term has a sign "Minus;" if in state - positive sign. Each term is equal to the product of the probability density of transition, which corresponds to this arrow/pointer, multiplied to the probability of that state from which proceeds the arrow/pointer.

This rule of the composition of differential equations for the probabilities of states is common/general/total and it is correct for any continuous Markov chain; with its aid it is possible completely mechanically, without any reasonings, to record/write the differential equations for the probabilities of states it is direct on the labeled graph/count of states.

Example. The labeled graph/count of the states of system S takes the form shown on Fig. 4.14. To write the system of the differential equations of Kolmogorov and initial conditions for the solution of this system, if it is known that at the initial moment the system is in state S_1.

Solution. The system of equations of Kolmogorov takes the form:

$$\frac{dp_1}{dt}=-(\lambda_{12}+\lambda_{13})\,p_1,$$

$$\frac{dp_2}{dt}=\lambda_{12}\,p_1+\lambda_{32}\,p_3,$$

$$\frac{dp_3}{dt}=-(\lambda_{32}+\lambda_{34})\,p_3+\lambda_{13}\,p_1+\lambda_{53}\,p_5,$$

$$\frac{dp_4}{dt}=-\lambda_{45}\,p_4+\lambda_{34}\,p_3,$$

$$\frac{dp_5}{dt}=-\lambda_{53}\,p_5+\lambda_{45}\,p_4.$$

Initial conditions:

with $t = 0$, $p_1 = 1$, $p_2 = p_3 = p_4 = p_5 = 0$.

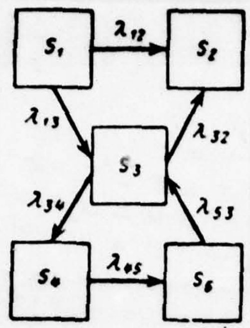


Fig. 4.14.

Page 200.

4. Flow of events the simplest flow and of its property.

In the examination of the random processes, which take place in systems with discrete states and continuous time, frequently it is necessary to meet the so-called "flows of events".

The flow of events is called the sequence of the uniform events, following one after another into some, generally speaking, random moments of time.

Examples they can be:

- flow of calls at exchange;

- the flow of the inclusions of instruments in everyday electric system;

- the flow of the cargo compositions, which enter railroad station;

- the flow of the malfunctions (short duration failures) of computer;

- the flow of the shots, directed to target/purpose and, etc.

In the examination of the processes, which take place in system with discrete states and continuous time, frequently there is to convenient visualize the process in the manner that as if the transitions of system from state into state occur under the action of some flows of events (flow of calls, the flow of malfunctions, the flew of claims for maintenance, the flow of visitors, etc.). Therefore has sense to examine in more detail the flows of events and their property.

Let us represent the flow of events as the sequence of points on time axis Ot (Fig. 4.15). Using this image, it is not necessary to forget, that the position of each point on the axis of abscissas is random.

The flow of events is called regular, if events follow one another through strict specific interval of time. This flow comparatively rarely is encountered in practice, but represents a definite interest as limiting case.

During operations research, more frequently it is necessary to meet the flows of the events, for which and the torque/moments of occurrence of an event and time intervals between them are random.

In this paragraph we will consider the flows of events, which possess some especially simple properties. For this, let us introduce a series of determinations.

1. Flow of events is called stationary, if hit probability of one or the other number of events to section of time of long τ (Fig. 4.15) depends only on length of section and does not depend upon where precisely on axis Ot is arrange/located this section.

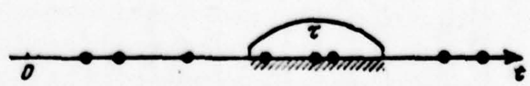


Fig. 4.15.

Page 201.

2. Flow of events is called flow without aftereffect, if for any nonintersecting sections of time number of events, which fall to one of them, does not depend on that, how many events hit to another (or others, if is examined more than two secticns).

3. Flow of events is called ordinary, if hit probability to elementary section of two or more events is negligible in comparison with hit probability of one event.

Let us examine in more detail these three properties of flows and will look, to which physical conditions they correspond and because of what can be broken.

The stability of flow indicates its uniformity on the time: the probabilistic characteristics of this flow must not vary depending on time. In particular, the so-called intensity (or "density") of the flow of events - the average number of events per unit time - for a stationary flow must remain constant. This, it goes without saying, not that means that the actual number of events, which appear per unit time, constantly - no, flow can have local condensation and resolutions. It is important that for a stationary flow these condensation and evacuation/rarefactions. It is important that for a

stationary flow these condensation and evacuation/rarefactions do not bear regular character, but the average number of events, which fall to the single section of time, remains constant for entire considered period.

In practice frequently are encountered the flows of the events which (at least, on the limited section of time) can be considered as stationary. For example, the flow of the calls, which enter exchange, let us say, that in range from 12 to 13 hours, it can be considered stationary. The same flow during whole days will be no longer stationary (by night the intensity of flow of calls is much less than in the daytime). Let us note that so Ja is the matter also with the majority of the physical processes which we call "stationary" - in actuality they are stationary only on the limited section of time, and the propagation of this section to infinity - only the convenient method, used for the purpose of simplification.

The absence of aftereffect in flow means that the events, which form flow, appear at successive moments of time independently of each other. For example, the flow of the passengers, entering to metro station, can be considered flow without aftereffect, because the reasons, which stipulated separate passenger's arrival precisely at given torque/moment, but not into other, as a rule, are not connected with analogous reasons for other passengers. If this dependence

appears, the condition of the absence of aftereffect proves to be broken.

Let us consider, for example, the flow of the freight trains, which go over the siding. If, according to safety conditions, they cannot follow one another more frequently than through the time interval τ_0, then between events in flow there is dependence, and the condition of the absence of aftereffect is broken. If interval τ_0 is small in comparison with the average/mean interval between trains $\bar{\tau}$, this disturbance/breakdown is unessential, but if interval τ_0 we compare with $\bar{\tau}$, it it is necessary to consider.

Page 202.

The ordinariness of flow means that the events in flow come one by one, but not by pairs, by sets of three, etc. For example, the flow of the clients, who are directed for barbershop, virtually can be considered as ordinary, what it is cannot about the flow of the clients, directed to ZAGS [civil registry office] for recording the reject. The flow of the attacks of destroyers on bomber, which is located above enemy territory, is ordinary, if they attack target/purpose one by one, and it is not ordinary, if they go into attack in pairs or sets of three.

If in the nonordinary flow of event they occur only by pairs, only by sets of three and sc forth, then can be it considered as ordinary "flow vapor", the "flow of sets of three", etc. Somewhat more complexly is matter, if the number of events, which form "package" (group of the simultaneously incoming events) is random. Then it is necessary together with the flow of packages to examine random variable X - number of events in package, and the mathematical model of flow becomes falser: it represents by itself not only sequence of the onsets of packages, but also the sequence of random variables - numbers of events in each package (Fig. 4.16), where x_1, x_2, ..., x_i, ... - value, taken by random variable X in the first, second and so forth packages. An example of the nonordinary flow of events with the random number of events in package - flow of coaches, which arrive to the sorting station ("package" is train).

Let us consider the flow of events, which possesses all three properties: stationary, without aftereffect, ordinary. This flow is called simplest (or stationary Poisson) flow. Name "simplest" is connected with the fact that the mathematical description of the events, connected with the simplest flows, proves to be simplest. Let us note that by the way that "idle time itself", it is on first glance, regular flow from strict constant intervals between events, is not in any way "protozoan" in the sense of the word named above: it possesses pronounced aftereffect, since the onsets of events are

connected by rigid functional dependence. Specifically, due to this aftereffect the analysis of the processes, connected with regular flows, proves to be, as a rule, it is more difficult, but not more easily in comparison with protozoa.

The simplest flow plays among other flows special role. Namely, it is possible to demonstrate which with superposition (mutual imposition) is sufficient large number of flows, which possess the aftereffect (provided they were stationary and were ordinary), is formed the total flow, which can be considered simplest, and thereby it is more precise, than the larger number of flows store/adds up[1].

FOOTNOTE [1]. For this, it is additionally required so that the store/added up flows would be congruent in intensity, i.e., so that among them it would not be, let us say, that one, that exceeds in intensity the sum of all others. ENDFOOTNOTE.

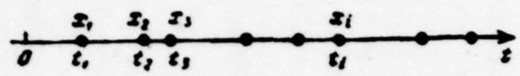


Fig. 4.16.

Page 203.

If the flow of events does not have an aftereffect, it is ordinary, but it is not stationary, it is called unsteady Poisson flow. In this flow intensity λ (average number of events per unit time) depends on the time:

$$\lambda = \lambda(t),$$

whereas for the simplest flow

$$\lambda = \text{const}.$$

The Poisson flow of events (both stationary and unsteady) is closely related with the known Poisson distribution. Namely, the number of events of flow, which fall to any section, is distributed according to the law of Poisson.

Let us explain, that this indicates. We will consider on axis

Ot, where is observed the flow of events, certain section of the time of length τ (see Fig. 4-17), that begins at torque/moment t_0 and finishing at torque/moment $t_0 + \tau$. It is not difficult to demonstrate (proof is given in all courses of the probability theory) that the hit probability to this section uniform t events is expressed by the formula:

$$P_m = \frac{a^m}{m!} e^{-a} \quad (m = 0, 1, \ldots), \qquad (4.1)$$

where a - average number of events, which is necessary to section τ.

For a stationary (simplest) Poisson flow value a is equal to the intensity of flow, multiplied by the length of the interval:

$$a = \lambda\tau,$$

i.e. on it depends upon where on axis Ot is undertaken section τ. For an unsteady Poisson flow value a is expressed by the formula:

$$a = \int_{t_0}^{t_0+\tau} \lambda(t)\, dt,$$

and, which means, that it depends on that, at which point t_0 begins the section τ.

Let us consider on axis Ot the simplest flow of events with intensity λ (Fig. 4.18).

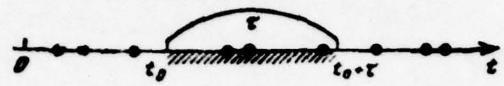


Fig. 4.17.

Page 204.

Us it will interest time interval T between adjacent events in this flow. It is obvious, T to eat a value is random; let us find its law of distribution. Let us first find the distribution function:

$$F(t)=P(T<t),$$

i.e. probability that value T will take value is less than t. Let us plot from the beginning of interval T (point t_0) segment t and will find probability that interval T will be lesser than t. For this, it is necessary that to the section of length t, which adjoins point t_0, would hit at least one the event of flow. Let us compute probability of this F(t) through the probability of the opposite event (to section t will hit not one events of flow):

$$F(t)=1-P_0.$$

Probability P_0 let us find from formula (1.4), set/assuming $m=0$:

$$P_0=\frac{a^0}{0!}e^{-a}=e^{-a}=e^{-\lambda t},$$

whence the distribution function of value T it will be:

$$F(t)=1-e^{-\lambda t}\quad (t>0). \qquad (4.2)$$

In order to find the density of distribution f(t) of random variable T, let us differentiate expression (4.2) on t:

$$f(t)=\lambda e^{-\lambda t}\quad (t>0). \qquad (4.3)$$

The law of distribution with density (4.3) is called exponential (or exponential). Its graph takes the form, presented in Fig. 4.19. Value λ is called the parameter of exponential law.

The exponential law of distribution, as we will see further, plays large role in the theory of the Markovian processes.

Let us find the numerical characteristics of random variable T - mathematical expectation (average value) m_t and dispersion D_t. We have:

$$m_t=\int_0^\infty tf(t)\,dt=\lambda\int_0^\infty te^{-\lambda t}\,dt.$$

Integrating in parts, we will obtain:

$$m_t=\frac{1}{\lambda}. \qquad (4.4)$$

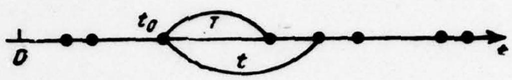


Fig. 4.18.

Page 205.

The dispersion of value T let us find through the second initial torque/moment:

$$D_t = \int_0^\infty t^2 f(t)\,dt - m_t^2 = \int_0^\infty \lambda t^2 e^{-\lambda t}\,dt - \frac{1}{\lambda^2},$$

whence, again integrating in parts, we will obtain:

$$D_t = \frac{1}{\lambda^2}. \qquad (4.5)$$

Taking the root square from dispersion, let us find the root-mean-square deviation of random variable T:

$$\sigma_t = \sqrt{D_t} = \frac{1}{\lambda}. \qquad (4.6)$$

Thus, for exponential distribution mathematical expectation and root-mean-square deviation are equal to each other and are reverse to the parameter λ.

It is concealed by form, tracing the structure of the simplest flow of events, we they arrived at the conclusion: time interval T between adjacent events in the simplest flow was distributed according to exponential law; its average value and root-mean-square deviation were equal to $1/\lambda$, where λ - an intensity of flow.

For unsteady Poisson flow the law of the distribution of interval/gap T will be no longer exponential; the form of this law will depend, in the first place, from that, where on axis 0t is arrange/located the first of the events, and, in the second place, from the form of dependence $\lambda(t)$, that characterizes alternating/variable intensity of flow. However, if $\lambda(t)$ varies comparatively slowly and its change for time between two events is small, then the law of the distribution of time interval between events it is possible to approximately count index (4.3), set/assuming in this formula value λ equal to average value $\lambda(t)$ on that section which us interests.

In conclusion of this paragraph let us deduce expression for the so-called "probability element of appearing the event".

Let us consider on axis 0t the simplest flow of events with intensity λ and the elementary section Δt, adjacent at point t (Fig. 4.20).

Let us find probability that on section Δt will appear some event of flow, i.e., section will not be "empty". Since flow is ordinary, the probability of appearance on section Δt more than one event can be disregarded.

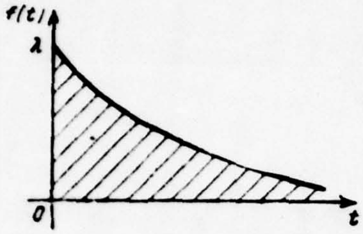


Fig. 4.19.

Fig. 4.20.

Page 206.

Let us designate $P_0(\Delta t)$ probability that on section Δt there will not be the events, but $P_1(\Delta t)$ - probability that on it will appear one event. On the basis force of the ordinariness of the flow

$$P_1(\Delta t) \approx 1 - P_0(\Delta t),$$

and probability $P_0(\Delta t)$ we compute on formula (4.1):

$$P_0(\Delta t) = \frac{a^0}{0!} e^{-a} = e^{-a} = e^{-\lambda \Delta t},$$

whence

$$P_1(\Delta t) \approx 1 - e^{-\lambda \Delta t}.$$

Expanding $e^{-\lambda \Delta t}$ in a series according to degrees $\lambda \Delta t$ and disregarding the small higher-order quantities, we will obtain:

$$P_1(\Delta t) \approx 1 - (1 - \lambda \Delta t).$$

Hence

$$P_1(\Delta t) \approx \lambda\Delta t, \qquad (4.7)$$

i.e. the probability of appearance on the elementary section of the time Δt of some event of flow is approximately equal to λΔt, where λ - an intensity of flow. This probability we will call the "probability element of appearing the event".

It is obvious, the same formula will be valid also for an unsteady Poisson flow, with that difference, that value λ must be taken equal to its value at that point t, which borders on the section Δt:

$$P_1(\Delta t) \approx \lambda(t)\,\Delta t.$$

5. Flows palm. Erlang's flows.

The flow of events is called flow palm (or by flow with the limited aftereffect), if time intervals between the consecutive events:

$$T_1,\ T_2,\ \ldots,\ T_i,\ \ldots$$

represent by themselves independent, equally distributed random variables (Fig. 4.21).

The simplest flow there is a special case of flow the palm: in

it distances T_1, T_2, ..., T_i, ... represent by themselves random variables, distributed according to one and that not exponential law; their independence follows from the fact that the simplest flow is a flow without aftereffect, and distance on time between any two events does not depend on are such the distances between each other.

Let us consider an example of flow palm. Certain cell/element of technical equipment/device (for example, electron tube) works continuously to its failure (breakdown), after which it is instantly substituted new. The life of cell/element is accidental.

Page 207.

If the separate copies of cell/elements go out of order independently of each other, then is the flow of failures (or the "flow is restored," since failure and restoration/reduction they occur at one and the same torque/moment) it represents by itself flow palm. If moreover life of cell/element is distributed according to exponential law, flow palm is converted into the simplest (stationary Poisson) flow.

Another example: the group of aircraft goes in combat formation "column" (Fig. 4.22) with an identical for all aircraft speed of V. Each of them, except that lead, is due to withstand system, i.e., to

be held at the assigned distance from in front of that go. This distance, measured by range finder, it is withstood with errors. The torque/moments of the intersection with the aircraft of the assigned border under these conditions form flow palm, since random variables $T_1 = L_1/V$; $T_2 = L_2/V$; ... are independent. Let us note that the same flow will not be flow the palm, if aircraft attempt to maintain the assigned distance not from soseda, but from which leads entire column.

Many flows of events, which are encountered in practice, although they are not in accuracy flows palm, can be by them approximately replaced.

The important for practice specimen/samples of flows palm are Erlang's so-called flows. These flows are formed as a result of the "sifting" of the simplest flows.

Let us consider on axis Ot of simplest flow of events (Fig. 4.23) and will preserve in it not all points, but only each the second; the others let us reject (in Fig. 4.23) the preserved points are shown heavily). As a result of this operation "of cutting" or "sifting" is formed again the flow of events; it is called the flow of Erlang of second order.

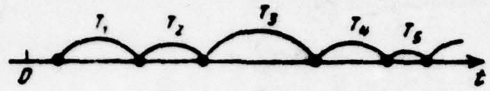


Fig. 4.21.

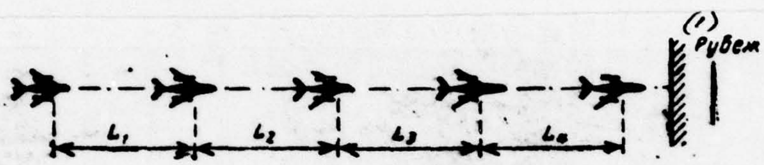


Fig. 4.22.

Key: (1). Border.

Fig. 4.23.

Page 208.

Generally, the flow of Erlang of k order $Э_k$ is called flow, it

is obtained, if we in the simplest flow preserve each k-th point, and the others to reject.

For example, Fig. 4.24 shows the formation/education of Erlang's flow of the 4th order $Э_4$ (three points of the simplest flow are rejected, and the fourth is retained).

It is obvious, the simplest flow represents by itself a special case of Erlang's flow, and precisely the flow of Erlang of 1st order $Э_1$.

The period of time T between adjacent events in Erlang's flow of the k order represents by itself sum k of independent random quantities - distances between events of the initial simplest flow:

$$T = T_1 + T_2 + \ldots + T_k = \sum_{i=1}^{k} T_i.$$

Each of these random variables is distributed according to the exponential law:

$$f_1(t) = \lambda e^{-\lambda t} \quad (t > 0).$$

The law of the distribution of interval of T between adjacent events in flow $Э_k$ is called the law of Erlang of k order.

Let us find expression for the density of distribution of this law; let us designate it $f_k(t)$. For this, let us consider on axis (Fig. 4.25) the simplest flow with intensity λ, in which the events are

divided by intervals T_1, T_2, ..., and let us find probability element $f_k(t)dt$ - probability that interval $T=\sum_{i=1}^{k}T_i$ will render/show within the limits of elementary section (t, t + dt).

For this, in the first place, to section with a length of t must hit exactly k - 1 points of the simplest flow; the probability of this event, according to formula (4.1), is equal to

$$P_{k-1}=\frac{a^{k-1}}{(k-1)!}e^{-a}=\frac{(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}.$$

Furthermore, last/latter (the k-th) point must hit to elementary section (t, t + dt) - the probability of this is equal to λdt (see formula (4.7)). Multiplying these probabilities, we will obtain:

$$f_k(t)\,dt=\frac{(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}\lambda\,dt,$$

whence

$$f_k(t)=\frac{\lambda(\lambda t)^{k-1}}{(k-1)!}e^{-\lambda t}\quad(t>0). \tag{5.1}$$

Fig. 4.24.

Page 209.

It is obvious, with k = 1, is obtained the usual exponential distribution:

$$f_1(t) = \lambda e^{-\lambda t} \quad (t > 0). \tag{5.2}$$

Let us find the characteristics of Erlang's law of the k order: his mathematical expectation $m_t^{(k)}$ and dispersion $D_t^{(k)}$. Random variable T, distributed according to the law of Erlang of k order, is obtained by addition k of the independent random quantities:

$$T = \sum_{i=1}^{k} T_i,$$

where each of values T_i is distributed according to exponential law (5.2) with mathematical expectation $1/\lambda$ and dispersion $1/\lambda^2$ (see formulas (4.4) and (4.5)). Applying the theorems of the addition of mathematical expectations and dispersions, we have

$$m_t^{(k)} = \frac{k}{\lambda}, \quad D_t^{(k)} = \frac{k}{\lambda^2}. \tag{5.3}$$

Extracting from last/latter expression square root, let us find the root-mean-square deviation:

$$\sigma_t^{(k)} = \frac{\sqrt{k}}{\lambda}.$$

Thus, we found mathematical expectation, dispersion and the root-mean-square deviation of the interval between adjacent events in Erlang's flow of k order:

$$m_t^{(k)} = \frac{k}{\lambda}; \quad D_t^{(k)} = \frac{k}{\lambda^2}; \quad \sigma_t^{(k)} = \frac{\sqrt{k}}{\lambda}. \tag{5.4}$$

Let us note that both the law of distribution $f_k(t)$, and all the its characteristics are expressed not through the intensity of Erlang's very flow $Э_k$, but through the intensity λ of his generating simplest flow which underwent cutting through. It is of interest to express them through the intensity (average number of events per unit time) of Erlang's very flow $Э_k$. Let us designate Λ_k - an intensity of flow $Э_k$.

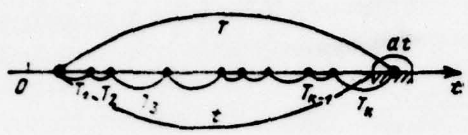


Fig. 4.25.

Page 210.

It is obvious

$$\Lambda_k = \lambda/k; \quad \lambda = k\Lambda_k,$$

since from initial simplest flow with intensity λ is taken only k part.

Substituting expression λ through Λ_k in formula (5.1), we will obtain

$$f_k(t) = \frac{k\Lambda_k (k\Lambda_k t)^{k-1}}{(k-1)!} e^{-k\Lambda_k t},$$

or

$$f_k(t) = \frac{(k\Lambda_k)^k}{(k-1)!} t^{k-1} e^{-k\Lambda_k t} \quad (t > 0). \tag{5.5}$$

Mathematical expectation, dispersion and the root-mean-square deviation of this law will be:

$$m_t^{(k)} = \frac{1}{\Lambda_k}; \quad D_t^{(k)} = \frac{1}{k\Lambda_k^2}; \quad \sigma_t^{(k)} = \frac{1}{\sqrt{k}\,\Lambda_k}. \tag{5.6}$$

Now let us assume that, retaining constant/invariable intensity of flow Λ_k:

$$\Lambda_k = \Lambda = \text{const},$$

we will vary only order k of Erlang's law. Its mathematical expectation will remain constant:

$$m_t = \frac{1}{\Lambda}, \qquad (5.7)$$

and dispersion and root-mean-square deviation will vary:

$$D_t^{(k)} = \frac{1}{k\Lambda^2}; \quad \sigma_t^{(k)} = \frac{1}{\sqrt{k}\Lambda}. \qquad (5.8)$$

From formulas (5.8) it is evident that with k $\rightarrow \infty$ and the dispersion, and root-mean-square deviation vanish. But that this does mean? This means that with k $\rightarrow \infty$ the flow of Erlang of assigned intensity Λ unlimitedly they approach regular flow with the constant interval between the events:

$$T = \text{const} = \frac{1}{\Lambda}.$$

This property of Erlang's flows conveniently in practical applications makes it possible, being assigned by different k, to obtain the flows, which possess different aftereffect - from the full/total/complete absence of aftereffect (k = 1) to the rigid functional connection between the onsets of events (k = ∞). Thus, the order of Erlang's flow k can serve in some degree as the "measure aftereffect".

For the purpose of simplification it is frequently to convenient approximately replace the real flow of events - by Erlang's flow with the same aftereffect. This are made, matching the characteristics of real flow - mathematical expectation and the dispersion of the interval between events - with the same characteristics of Erlang's substituting flow.

Page 211.

Example. As a result of statistical processing of time intervals between events in certain flow are obtained the following characteristics:

- average value of interval $m_t = 2$ min,

- the root-mean-square deviation of interval $\sigma_t = 0{,}9$ min.

It is required to select Erlang's flow, which possesses approximately the same characteristics, to find his intensity Λ and order k.

Solution. Intensity Λ there is the value, reciprocal to the

average interval between the events:

$$\Lambda = 1/m_t = 1/2 = 0{,}5$$ (event/min.).

From formula (5.8) we find the order of Erlang's flow k:

$$k = \left(\frac{1}{\sigma_t \Lambda}\right)^2 = \left(\frac{1}{0{,}9 \cdot 0{,}5}\right)^2 \approx 4{,}9.$$

Choosing as k near integer we obtain

$$k = 5.$$

Thus, this flow can be approximately replaced with the flow of Erlang of 5th order with a density of the form of:

$$f_5(t) = \frac{(5 \cdot 0{,}5)^5}{4!} t^4 e^{-5 \cdot 0{,}5 t}$$

or

$$f_5(t) = 4{,}1 t^4 e^{-2{,}5 t} \quad (t > 0). \qquad (5.9)$$

The form of the curve of distribution (5.9) is shown on Fig. 4.26).

The special attention, given here to the flows of Erlang in comparison with other flows palm (with the arbitrary law of timing between adjacent events) is explained by the fact that with the help of these flows it is possible to reduce non-Markov processes to Markov. As this is made, we will see later, in §§10, 11 of present

chapter, and also in §6 of chapter 5.

By Erlang's flow are very convenient for the approximate representation of flows the palm of any form, since the flows of Erlang of different orders form whole gamma, which gives gradual transition from the simplest flow (full/total/complete absence of aftereffect) to flow with regular intervals (full/total/complete, rigid aftereffect). The possibility of the approximate representation of any flows palm by flows of Erlang's type they are even more widened, if we use "generalized Erlang's laws", which they are obtained during the addition of several random variables, distributed according to exponential laws with the different parameters (for example, see [8]), and Erlang's also "mixed generalized laws", which they are obtained, if we sum several Erlang's generalized laws with the coefficients ("weights"), which form in sum one.

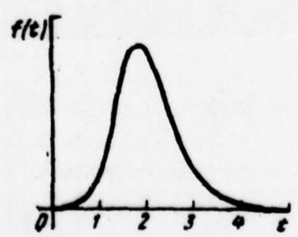


Fig. 4.26.

6. Poisson flows of events and continuous Markov chains.

Page 212.

Let us consider certain physical system S with the discrete states S_1, S_2, ..., S_n which passes from state into state under the effect of some random events, for example, calls at exchange, breakdowns (failures) of the cell/elements of apparatus, the shots, directed along target/purpose, etc.

Let us this visualize so, as if the events, which translate system from state into state, represent by themselves some flows of events (flows of calls, the flows of failures, the flows of shots, etc.).

Let system S with the graph/count of states, shown on Fig. 4.27, at torque/moment t is be in state S_i and can pass from it into state S_j under the effect of some Poisson flow of events with an intensity of λ_{ij}: as soon as it appears the first event of this flow system instantly passes (it jumps) from S_i in S_j. As we know that this

transitional probability for elementary time interval Δt (probability element of transition) is equal to $\lambda_{ij}\Delta t$. Thus, the probability density of transition λ_{ij} in continuous Markov chain represents by itself nothing else but the intensity of flow of events, which translates system on the appropriate arrow/pointer.

If all flows of events, which translate system S from state into state, Poisson (stationary or unsteady - are unimportant), then the process, which takes place in system, will be Markov. It is real/actual, Poisson flow possesses the absence of aftereffect, therefore, in the assigned state of system at given torque/moment, its transitions into other states in the future are caused only by the appearance of some events in Poisson flows, but the probabilities of appearing these events do not depend on the "prehistory" of process.

In the future, examining Markov processes in systems with discrete states and continuous time (continuous Markov chains), to us is convenient will be in all cases to consider the transitions of system from state into state as occurring under the effect of some flows of events, at least in actuality these events were single. For example, the working technical equipment/device we will consider as being located under the action of the flow of failures, although actually it can refuse only one time. It is real/actual, if

equipment/device it rejects at that torque/moment when comes the first event of flow, then completely nevertheless - is continued after this the flow of failures or it ceases: the fate of equipment/device on this no longer depends. For us it will more conveniently deal precisely with the flows of events.

Thus, is examined system S, in which the transitions from state into state occur under the action of the Poisson flows of events with the specific intensities. Let us write these intensities (probability density of transitions) on the graph/count of the states of system of appropriate rifleman/gunner. We will obtain the labeled graph/count of states (Fig. 4.27); on which, using the rule, formulated into §3, it is possible to immediately register the differential equations of Kolmogorov for the probabilities of states.

Page 213.

Example 1. The technical system S consists of two assemblies: I and II; each of them independent of other can reject (go out of order). The flow of the failures of the first assembly - Poisson, with intensity λ_I; the second - also Poisson, with intensity λ_{II}. Each assembly right after failure begins to be overhauled (to be restored). Flow of restoration/reductions (terminations of the repair of the overhauled assembly) for both assemblies - Poisson with intensity λ.

To comprise the graph/count of the states of system and to write the equations of Kolmogorov for the probabilities of states. To determine, under which initial conditions it is necessary to solve these equations, if at the initial moment (t = 0) system works exactly.

Solution. States of the system:

S_{11} - both assembly are exact,

S_{21} - first assembly it is overhauled, the second is exact,

S_{12} - the first assembly it is exact, the second is overhauled,

S_{22} - both assembly are overhauled.

The labeled graph/count of the states of system is shown on Fig. 4.28.

The intensities of flow of events on Fig. 4.28 are written from following considerations. If system S is in state S_{11}, then on it function two flows of events the flow of the malfunctions of

node/unit I with intensity λ_{II}, which translates it into state S_{21}, and the flow of the malfunctions of node/unit II with intensity that translates it into S_{12}. Let now the system is be in state S_{21} (node/unit I is overhauled, node/unit II is exact). From this state the system can, in the first place, return to S_{11} (this occurs under the action of the flow of restoration/reductions with intensity λ); in the second place, - to pass into state S_{22} (when the repair of node/unit I is not still finished, but node/unit II meanwhile left the system); this transition occurs under the action of the flow of the failures of node/unit II with intensity λ_{II}. Intensities of flow at remaining rifleman/gunner are enter/written analogously.

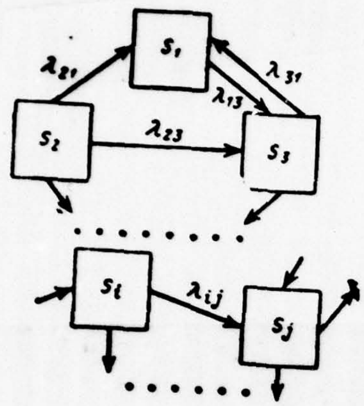


Fig. 4.27.

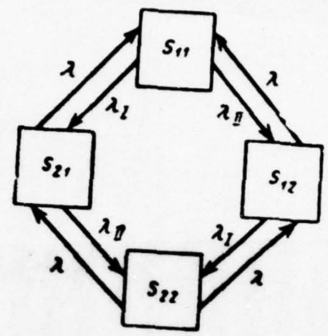


Fig. 4.28.

Page 214.

Designating the probabilities of states p_{11}, p_{21}, p_{12} and p_{22} and by using the rule, formulated into §3, let us register the equations of Kolmogorov for the probabilities of the states:

$$
\left.\begin{aligned}
\frac{dp_{11}}{dt} &= -(\lambda_I+\lambda_{II})\,p_{11}+\lambda p_{21}+\lambda p_{12},\\
\frac{dp_{21}}{dt} &= -(\lambda+\lambda_{II})\,p_{21}+\lambda_I\,p_{11}+\lambda p_{22},\\
\frac{dp_{12}}{dt} &= -(\lambda+\lambda_I)\,p_{12}+\lambda_{II}\,p_{11}+\lambda p_{22},\\
\frac{dp_{22}}{dt} &= -2\lambda p_{22}+\lambda_{II}\,p_{21}+\lambda_I\,p_{12}.
\end{aligned}\right\} \quad (6.1)
$$

The initial conditions under which it is necessary to solve this

system:

with $t = 0$ $p_{11} = 1$, $p_{21} = p_{12} = p_{22} = 0$.

Let us note that, using the condition

$$p_{11} + p_{21} + p_{12} + p_{22} = 1,$$

it would be possible to decrease the number of equations on one. It is real/actual, any of the probabilities p_{11}, p_{21}, p_{12}, p_{22} can be expressed by the others and substituted into equations (6.1), but the equation, which contains on the left side the derivative of this probability of - reject/throwing.

Let us note that besides the fact that equations (6.1) are valid both for the constant intensities of Poisson flows $\lambda_I, \lambda_{II}, \lambda,$ and for variables

$$\lambda_I = \lambda_I(t); \quad \lambda_{II} = \lambda_{II}(t); \quad \lambda = \lambda(t).$$

Example 2. Group in the composition of five aircraft in line astern (Fig. 4.29) accomplishes coating on the territory of enemy. Front/leading aircraft (leading) is jammer; until it is biased/beaten, the following it aircraft cannot be discovered and are attacked by the air defense weapons of enemy. Attacks undergoes only jammer. Flow of attacks - poisson, with intensity λ (attacks/hour).

As a result of attack jammer is surprised with probability p.

If jammer is struck (it is biased/beaten), then following after it aircraft are discovered and undergo attacks PVO; for each of them (until it is struck) it is directed the Poisson flow of attacks with intensity λ; by each attack aircraft is surprised with probability p. When aircraft is struck, attacks on it cease, but to other aircraft they are not transferred.

To write the equations of Kolmogorov for the probabilities of the states of system and to indicate initial conditions.

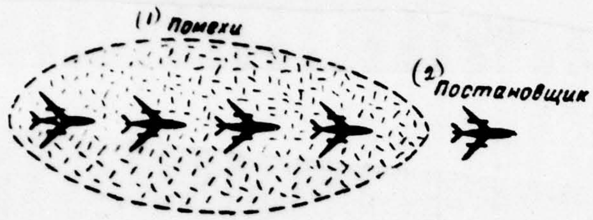


Fig. 4.29.

Key: (1). Interferences. (2). Producer.

Page 215.

Solution. Let us label the states of system with respect to the number preserved aircraft in the group:

S_5 - all aircraft are whole;

S_4 - jammer is biased/beaten, remaining aircraft whole;

S_3 - jammer and one bomber are biased/beaten, remaining aircraft whole;

S_2 - jammer and two bombers are biased/beaten, remaining aircraft whole;

S_1 - jammer and three bombers are biased/beaten, one aircraft is whole;

S_0 - all aircraft are biased/beaten.

States we differ from each other by the number of preserved bombers, but not on that, which precisely of them was preserved, since all bombers according to the conditions of problem were equivalent - they attack with identical intensity they are surprised with identical probability.

The graph/count of the states of system is shown on Fig. 4.30. In order to label this graph/count, let us determine the intensities of flow of the events, which translate system from state into state.

From state S_5 into S_4 the system translates the flow of the damaging (or "successful") attacks, i.e., those attacks which lead to the defeat of producer (it goes without saying that if it was not earlier struck). The intensity of flow of attacks is equal to λ, but not all they - damaging: each of them proves to be damaging only with probability p. It is obvious, intensity of the flow of the damaging

attacks is equal to λp; this intensity is written as λ_{54} of the first to the left arrow/pointer on graph/count (Fig. 4.30).

We will be occupied following arrow/pointer and will find intensity λ_{43}. System is in state S_4, i.e., are whole and can be attacked four aircraft. It will pass into state S_3 for time Δt, if for this time of any of the aircraft (nevertheless, which) will be biased/beaten. Let us find the probability of opposite event - in time Δt not one aircraft it will not be biased/beaten:

$$(1-\lambda p\Delta t)(1-\lambda p\Delta t)(1-\lambda p\,\Delta t)(1-\lambda p\,\Delta t)=(1-\lambda p\,\Delta t)^4\approx 1-4\lambda p\,\Delta t.$$

Are here reject/thrown the members of the higher order of smallness relative to Δt. Subtracting this probability from unit, we will obtain transitional probability from S_4 in S_3 for time Δt (probability element of transition):

$$4\lambda p\Delta t,$$

whence

$$\lambda_{43}=4\lambda p,$$

that also is written of the second to the left arrow/pointer. Let us note that the intensity of this flow of events is simply equal to the sum of the intensities of flow of the damaging attacks, directed toward separate aircraft. Discussing visually, it is possible to

AD-A066 460 FOREIGN TECHNOLOGY DIV WRIGHT-PATTERSON AFB OHIO F/G 12/2
OPERATIONS RESEARCH. PART I,(U)
JUN 78 Y S VENTTSEL'
UNCLASSIFIED FTD-ID(RS)T-0687-78-PT-1 NL

6 OF 6
AD
A066460

END
DATE
FILMED
5-79
DDC

obtain this conclusion/derivation as follows: system S of state S_4 consists of four aircraft; for each of them, functions the flow of the damaging attacks with intensity λp; means to system as a whole it functions the total flow of the damaging attacks with intensity $4\lambda p$.

With the help of analogous reasonings are enter/written the intensities of flow of events at remaining rifleman/gunner.

$$\boxed{S_5} \xrightarrow{\lambda p} \boxed{S_4} \xrightarrow{4\lambda p} \boxed{S_3} \xrightarrow{3\lambda p} \boxed{S_2} \xrightarrow{2\lambda p} \boxed{S_1} \xrightarrow{\lambda p} \boxed{S_0}$$

Fig 4.30

Page 216.

The equations of Kolmogorov for the probabilities of states take the form:

$$\left.\begin{aligned}
\frac{dp_5}{dt} &= -\lambda p p_5, \\
\frac{dp_4}{dt} &= -4\lambda p p_4 + \lambda p p_5, \\
\frac{dp_3}{dt} &= -3\lambda p p_3 + 4\lambda p p_4, \\
\frac{dp_2}{dt} &= -2\lambda p p_2 + 3\lambda p p_3, \\
\frac{dp_1}{dt} &= -\lambda p p_1 + 2\lambda p p_2, \\
\frac{dp_0}{dt} &= \lambda p p_1.
\end{aligned}\right\}$$

Since at the initial moment (with t = 0) all aircraft are whole, initial conditions will be:

$$\text{with } t=0 \quad p_5=1, \quad p_4=p_3=p_2=p_1=p_0=0.$$

Example [redacted] 3. The conditions the same as into example [redacted] 2, but intensity λ is related to the common/general/total flow of attacks, directed to entire group. While jammer is whole, all these attacks are directed for it; when it is biased/beaten, attacks are distributed evenly between the remaining aircraft, so that to one aircraft it comes on the average λ/k (attacks/hour), where k - a number of preserved aircraft. To comprise graph of states, to label it and to register the equations of Kolmogorov for the probabilities of states.

Solution. The labeled graph/count of states is shown on Fig. 4.31.

Equations of Kolmogorov!

$$\left.\begin{aligned}
\frac{dp_5}{dt} &= -\lambda p p_5,\\
\frac{dp_4}{dt} &= -\lambda p p_4 + \lambda p p_5,\\
\frac{dp_3}{dt} &= -\lambda p p_3 + \lambda p p_4,\\
\frac{dp_2}{dt} &= -\lambda p p_2 + \lambda p p_3,\\
\frac{dp_1}{dt} &= -\lambda p p_1 + \lambda p p_2,\\
\frac{dp_0}{dt} &= \lambda p p_1.
\end{aligned}\right\}$$

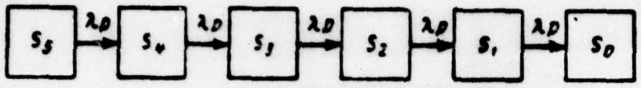


Fig 4.31

Page 217.

Let us note that in this paragraph we cnly wrote out differential equations for the probabilities of states, but they were not occupied by the solution of these equations.

In regard to this it is possible tc note following. Equations for the probabilities of states represent by themselves linear differential equations with constant or variable coefficients - depending on that, are constant or alternating/variable intensities λ_{ij} of the flows of the events, which translate system from state into state.

The system of several linear differential equations of such type only in rare cases can be integrated in the quadratures: usually this system it is necessary to solve numerically - either by hand or in analog computer (AVM), or finally by ETsVM. All these methods of the solution of the systems of the differential equations of difficulties do not supply/deliver; therefore the most essential - to be able to register system of equations and to formulate for it initial conditions, than we were bounded here.

7. Maximum probabilities of states.

Let there be the physical system S with the discrete states:

$$S_1, S_2, \ldots, S_n,$$

in which proceeds the Markovian process with continuous time (continuous Markov chain). The graph/count of states is shown on Fig. 4.32.

Let us assume that all the intensities of flow of the events, which translate system from state into state, are constant:

$$\lambda_{ij} = \text{const},$$

in other words, all flows of events - the simplest (stationary Poisson) flows.

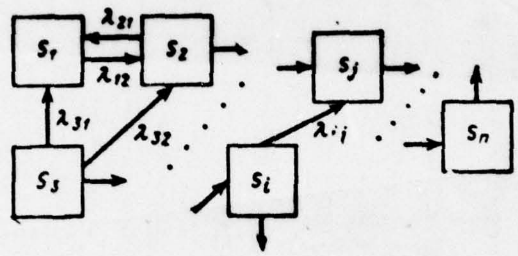


Fig. 4.32.

After registering the system of the differential equations of Kolmogorov for the probabilities of states and after integrating these equations under the assigned initial conditions, we will obtain the probabilities of states as functions of time, i.e., n of the functions:

$$p_1(t),\ p_2(t), \ldots,\ p_n(t),$$

with any t of those giving in sum one

$$\sum_{i=1}^{n} p_i(t) = 1.$$

Let us raise now the following question: that will occur with system S with $t \to \infty$? Will functions $p_1(t), \ldots, p_n(t)$ approach some limits? These limits, if they exist, are called the maximum (or "final") probabilities of states.

It is possible to demonstrate the following general consideration. If the number of states of system S certainly and from each state is possible to move (for one or the other number of step/pitches) into each another, then the maximum probabilities of states there exist and they do not depend on the initial state of system.

On Fig. 4.33, is shown the graph/count of states, who satisfies the placed condition: from any state the system can sooner or later pass into any another. On the contrary, for the system the graph/count of states of which is shown on Fig. 4.34, condition satisfied. It is obvious that if the initial state of such of systems S_1, then, for example, state S_6 with $t \to \infty$ can be reached, and if the initial state S_2 - cannot.

Let us assume that the placed condition is satisfied, and maximum probabilities exist:

$$\lim_{t\to\infty} p_i(t) = p_i \quad (i = 1, 2, \dots, n). \tag{7.1}$$

Maximum probabilities we will designate by the same letters p_1, p_2, ..., p_n, that also very probabilities of states, understanding raise this time variable values (function of time), but constant numbers.

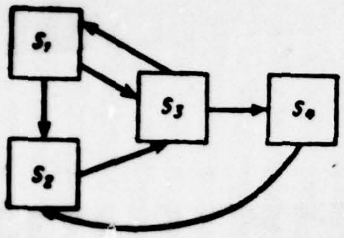


Fig. 4.33.

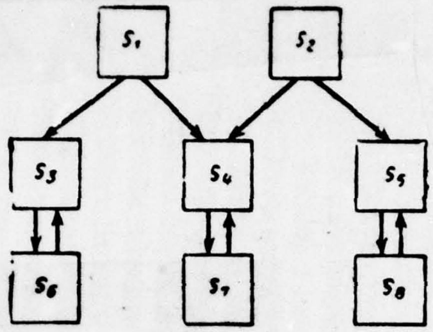


Fig. 4.34.

Page 219.

It is obvious, the maximum probabilities of states, just as pre-limit, in sum must give the unit:

$$\sum_{i=1}^{n} p_i = 1.$$

Thus, with $t \to$ - in system S is establish/installed not which maximum steady state: it lies in the fact that the system randomly varies its states, but the probability of each of them no longer depends on the time: each of the states is realized with certain constant probability. Which sense of this probability? It represents by itself nothing else but the mean relative retention time of system in this state. For example, if of system S three of the possible

state S_1, S_2 and S_3, moreover their maximum probabilities are equal to 0.2, 0.3 and 0.5, this means that after transition to the steady-state conditions/mode system S in of mean two tenth time will be located in state S_1, three tenth - in state S_2 and the half of time - in state S_3. Does arise the question: how to compute the maximum probabilities of states p_1, $p_2, \ldots, p_n$?

It turns out that for this in the system of equations of Kolmogorov, which describe the probabilities of states, it is necessary to assume all left sides (derivatives) equal to zero.

It is real/actual, in maximum (being steady) conditions/mode all probabilities of states are constant, which means, that their derivatives are equal to zero.

If all the left sides of the equations of Kolmogorov for the probabilities of states are assumed equal to zero, then the system of differential equations will be converted into the system of linear algebraic equations. Together with the condition

$$\sum_{i=1}^{n} p_i = 1 \qquad (7.2)$$

(the so-called "normalizing condition") these equations make it possible to compute all the maximum probabilities

Example 1. The physical system S has the possible states S_1, S_2, S_3, S_4 whose labeled graph/count is given on Fig. 4.35 (of each arrow/pointer placed numerical value of the corresponding intensity). To compute the maximum probabilities of states p_1, p_2, p_3, p_4.

Solution. We write the equations of Kolmogorov for the probabilities of the states

$$\left.\begin{aligned}
\frac{dp_1}{dt} &= -5p_1 + p_3,\\
\frac{dp_2}{dt} &= -p_2 + 2p_1 + 2p_3,\\
\frac{dp_3}{dt} &= -3p_3 + 3p_1 + 2p_4,\\
\frac{dp_4}{dt} &= -2p_4 + p_2.
\end{aligned}\right\} \qquad (7.3)$$

Page 220.

Set/assuming left sides equal to zero, we will obtain the system of algebraic equations for the maximum probabilities of the states:

$$\left.\begin{aligned}
0 &= -5p_1 + p_3,\\
0 &= -p_2 + 2p_1 + 2p_3,\\
0 &= -3p_3 + 3p_1 + 2p_4,\\
0 &= -2p_4 + p_2.
\end{aligned}\right\} \qquad (7.4)$$

Equations (7.4) - the so-called homogeneous equations (without absolute term). As is known from algebra, these equations determine values p_1, p_2, p_3, p_4 only with an accuracy to constant factor. Fortunately, of us exists the normalizing condition:

$$p_1 + p_2 + p_3 + p_4 = 1. \tag{7.5}$$

which, together with equations (7.4), it makes it possible to find all unknown probabilities.

It is real/actual, it is expressed from (7.4) all unknown probabilities through one of them, for example, through p_1. From the first equation:

$$p_3 = 5p_1.$$

Substituting in the second equation, we will obtain

$$p_2 = 2p_1 + 2p_3 = 2p_1 + 10p_1 = 12p_1.$$

Fourth equation gives

$$p_4 = {}^1/_2 p_2 = 6p_1.$$

Substituting all these expressions for p_2, p_3, p_4 under normalizing condition (7.5), we will obtain

$$p_1 + 12p_1 + 5p_1 + 6p_1 = 1.$$

Hence

$$24p_1 = 1, \quad p_1 = {}^1/_{24}, \quad p_2 = 12p_1 = {}^1/_2,$$
$$p_3 = 5p_1 = {}^5/_{24}, \quad p_4 = 6p_1 = {}^1/_4.$$

Thus, the maximum probabilities of states are obtained, they equal to

$$p_1 = {}^1/_{24}, \quad p_2 = {}^1/_2, \quad p_3 = {}^5/_{24}, \quad p_4 = {}^1/_4. \tag{7.6}$$

This means that in the maximum, steady-state conditions/mode system S will carry out able S_1 on the average one twenty fourth part of the time, able S_2 - half of time, able S_3 - five twenty fourth and able S_4 - one fourth of time.

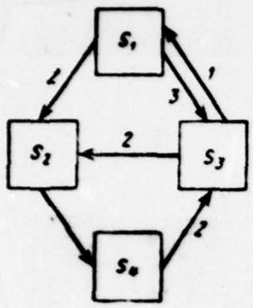


Fig. 4.35.

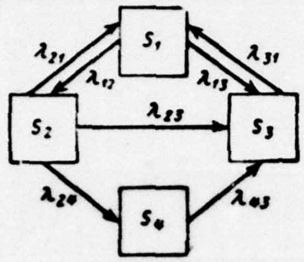


Fig. 4.36.

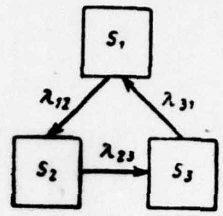


Fig. 4.37.

Page 221.

Let us note that solving this problem, we in no way used one of equations (7.4) - by the third. It is not difficult to ascertain that it is the corollary of three others: stcre/adding up all the four equations, we will obtain identical zero. With an equal success, solving system, we could reject/throw any of four equations (7.4).

The used by us method of the compcsition of algebraic equations for the maximum probabilities of states was reduced to following: to first write differential equations, and then to place in them left sides equal to zero. However, it is possible to register the algebraic equations for maximum probabilities and it is direct,

without passing through the stage cf differential. Let us illustrate this based on example.

Example 2. The graph/count of the states of system is shown on Fig. 4.36. To write algebraic equations for the maximum probabilities of states.

Solution. Without record/writing differential equations, we openly write the appropriate right sides and we equate to their zero; in order not to deal with negative terms, is immediately transferred to them another part, reversing the sign:

$$\left.\begin{aligned} &\lambda_{21} p_2 + \lambda_{31} p_3 = (\lambda_{12} + \lambda_{13}) p_1, \\ &\lambda_{12} p_1 = (\lambda_{23} + \lambda_{24}) p_2, \\ &\lambda_{13} p_1 + \lambda_{23} p_2 + \lambda_{43} p_4 = \lambda_{31} p_3, \\ &\lambda_{24} p_2 = \lambda_{43} p_4. \end{aligned}\right\} \tag{7.7}$$

In order subsequently to immediately write such equations, it is useful to memorize the following mnemonic rule: "which flows, then also it escape/ensues", i. e., for each state the sum of the terms, which correspond entering to arrow/pointers, it is equal to the sum of the terms, which correspond outgoing; each term is equal to the intensity of flow of events, which translates system on this arrow/pointer, multiplied by the probability of that state which leaves the arrow/pointer.

Example 3. To write algebraic equations for the maximum probabilities of the states of system S, the graph/count of states of which is given on Fig. 4.37. To solve these equations.

Solution. We write algebraic equations for the maximum probabilities of the states:

$$\left.\begin{aligned} \lambda_{31} p_3 &= \lambda_{12} p_1, \\ \lambda_{12} p_1 &= \lambda_{23} p_2, \\ \lambda_{23} p_2 &= \lambda_{31} p_3. \end{aligned}\right\} \qquad (7.8)$$

Normalizing condition:

$$p_1 + p_2 + p_3 = 1. \qquad (7.9)$$

Is expressed with the help of first two equations (7.8) p_2 and p_3 through p_1:

$$\left.\begin{aligned} p_3 &= \frac{\lambda_{12}}{\lambda_{31}} p_1; \\ p_2 &= \frac{\lambda_{12}}{\lambda_{23}} p_1. \end{aligned}\right\} \qquad (7.10)$$

Let us substitute into them normalizing condition (7.9)

$$p_1 + \frac{\lambda_{12}}{\lambda_{23}} p_1 + \frac{\lambda_{12}}{\lambda_{31}} p_1 = 1.$$

whence

$$p_1 = \frac{1}{1 + \frac{\lambda_{12}}{\lambda_{23}} + \frac{\lambda_{12}}{\lambda_{31}}}.$$

Further, from (7.10) we will obtain

$$p_2 = \frac{\frac{\lambda_{12}}{\lambda_{23}}}{1 + \frac{\lambda_{12}}{\lambda_{23}} + \frac{\lambda_{12}}{\lambda_{31}}}; \quad p_3 = \frac{\frac{\lambda_{12}}{\lambda_{31}}}{1 + \frac{\lambda_{12}}{\lambda_{23}} + \frac{\lambda_{12}}{\lambda_{31}}}.$$

Page 222.

8. Process of "death and of multiplication".

In the previous paragraph we ascertained that by knowing the labeled graph/count of the states of system, it is possible tc immediately write algebraic equations for the maximum probabilities of states. Thus, if two continuous chains. Markov have the identical graphs of states and are distinguished only by the values of intensities λ_{ij}, Then there is

no necessity to find the maximum probabilities of states for each of the graph/counts individually: sufficient to comprise and to solve for the literal form of equation for one of them, and then to substitute for λ_{ij}, the corresponding values. For graph/counts's many frequently encountered forms, linear equations easily are solved in literal form.

In this paragraph we will be introduced to one very typical pattern of continuous Markov chains - the so-called "set-up of death and multiplication" [1].

FOOTNOTE [1]. The origin of term "set-up of death and multiplication" originates from the biological problems where by a similar set-up is described the process of changing the number of population. ENDFOOTNOTE.

Markov continuous chain is called the "process of death and multiplication", if its graph/count of states takes the form, presented in Fig. 4.38, i.e., all states it is possible to draw out into one chain/network in which each of the average states (S_2, ..., S_{n-1}) is connected by direct/straight and feedback with each of the adjacent states, and end states (S_1, S_n) only with one adjacent state.

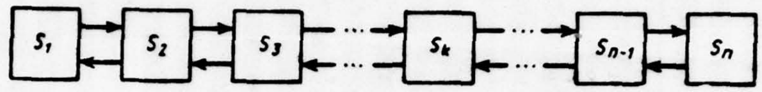


Fig. 4.38.

Page 223.

Example 1. Technical equipment/device consists of three identical assemblies; each of them can go cut cf order (reject); from the seemed assembly immediately begins to be restored. The states of system we label according to the number of defective assemblies:

S_0 - all three assemblies are exact;

S_1 - one assembly refused (it is restored), two corrected:

S_2 - two assemblies are restored, one it is exact;

S_3 - all three assemblies are restored.

The graph/count of states is shown on Fig. 4.39. From graph it

is evident that the process, which takes place in system, represents by itself the process of "death and multiplication".

The set-up of death and multiplication very frequently is encountered in the most diverse practical problems; therefore has sense to previously consider this set-up in general form and to solve the matching system of algebraic equations with the fact, in order subsequently, meeting the concrete/specific/actual processes, which take place according to this set-up, not to solve problem each time anew, but to use already prepared/finished solution.

Thus, let us consider the random process of death and multiplication with the graph/count of states, presented in Fig. 4.40.

Let us write algebraic equations for the probabilities of states. For the first state S_1, we have:

$$\lambda_{12} p_1 = \lambda_{21} p_2. \tag{8.1}$$

For the second state S_2 the sums of the terms, which correspond to the entering and outgoing arrow/pointers, are equal to:

$$\lambda_{23} p_2 + \lambda_{21} p_2 = \lambda_{12} p_1 + \lambda_{32} p_3.$$

But, by force (8.1), it is possible to shorten to the right and

to the left equal to each other members $\lambda_{12}p_1$ and $\lambda_{21}p_2$; we will obtain:

$$\lambda_{23}p_2=\lambda_{32}p_3,$$

and further, in perfect analogy,

$$\lambda_{34}p_3=\lambda_{43}p_4,$$

$$\cdots\cdots$$

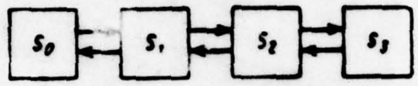


Fig. 4.39.

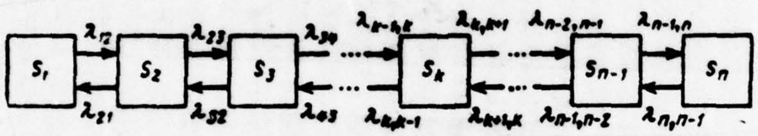


Fig. 4.40.

Page 224.

In a word, for the circuit of death and multiplication members, who correspond confronting above each other to arrow/pointers, are equal to each other:

$$\lambda_{k-1,\,k}\, p_{k-1} = \lambda_{k,\,k-1}\, p_k, \qquad (8.2)$$

where k takes all values from 2 to n.

Thus, the maximum probabilities of states $p_1, p_2, \ldots, p_n$ in any set-up of death and multiplication satisfy the equations:

$$\left.\begin{aligned} \lambda_{12} p_1 &= \lambda_{21} p_2, \\ \lambda_{23} p_2 &= \lambda_{32} p_3, \\ \lambda_{34} p_3 &= \lambda_{43} p_4, \\ &\cdots\cdots\cdots \\ \lambda_{k-1,\,k} p_{k-1} &= \lambda_{k,\,k-1} p_k, \\ &\cdots\cdots\cdots \\ \lambda_{n-1,\,n} p_{n-1} &= \lambda_{n,\,n-1} p_n \end{aligned}\right\} \tag{8.3}$$

and the normalizing condition:

$$p_1 + p_2 + \ldots + p_n = 1. \tag{8.4}$$

Let us solve this system as follows: from first equation (7.3) it is expressed p_2:

$$p_2 = \frac{\lambda_{12}}{\lambda_{21}} p_1, \tag{8.5}$$

from the second, taking into account (8.5), we will obtain:

$$p_3 = \frac{\lambda_{23}}{\lambda_{32}} p_2 = \frac{\lambda_{23}\lambda_{12}}{\lambda_{32}\lambda_{21}} p_1, \tag{8.6}$$

from the third, taking into account (8.6):

$$p_4 = \frac{\lambda_{34}\lambda_{23}\lambda_{12}}{\lambda_{43}\lambda_{32}\lambda_{21}} p_1,$$

and generally

$$p_k = \frac{\lambda_{k-1,\,k}\lambda_{k-2,\,k-1}\cdots\lambda_{12}}{\lambda_{k,\,k-1}\lambda_{k-1,\,k-2}\cdots\lambda_{21}} p_1. \tag{8.7}$$

This formula is valid for any k from 2 to n.

Let us focus attention on its structure. In numerator stands the product of all probability densities of transition (intensities) λ_{ij}, that stand of the arrow/pointers, directed from left to right, from beginning and up to that that goes into state S_k; in denominator - product of all intensities λ_{ij}, which stand of the arrow/pointers, which go from right to left, furthermore, from beginning and up to the arrow/pointer, which proceeds from state S_k. With k = n in numerator, it will stand the product of intensities λ_{ij}, which stand at all rifleman/gunner, that go from left to right, and in denominator - at all rifleman/gunner, that go from right to left.

Page 225.

Thus, all probabilities $p_1, p_2, \ldots, p_n$ are expressed through one of them: p_1. Let us substitute these expressions under the normalizing condition $p_1 + p_2 + \ldots + p_n = 1$. We will obtain

$$p_1+\frac{\lambda_{12}}{\lambda_{21}}p_1+\frac{\lambda_{23}\lambda_{12}}{\lambda_{32}\lambda_{21}}p_1+\ldots+\frac{\lambda_{k-1,\,k}\lambda_{k-2,\,k-1}\ldots\lambda_{12}}{\lambda_{k,\,k-1}\lambda_{k-1,\,k-2}\ldots\lambda_{21}}p_1+$$
$$+\ldots+\frac{\lambda_{n-1,\,n}\lambda_{n-2,\,n-1}\ldots\lambda_{12}}{\lambda_{n,\,n-1}\lambda_{n-1,\,n-2}\ldots\lambda_{21}}p_1=1,$$

whence

$$p_1=\frac{1}{1+\frac{\lambda_{12}}{\lambda_{21}}+\frac{\lambda_{23}\lambda_{12}}{\lambda_{32}\lambda_{21}}+\ldots+\frac{\lambda_{k-1,\,k}\lambda_{k-2,\,k-1}\ldots\lambda_{12}}{\lambda_{k,\,k-1}\lambda_{k-1,\,k-2}\ldots\lambda_{21}}+\ldots+\frac{\lambda_{n-1,\,n}\ldots\lambda_{12}}{\lambda_{n,\,n-1}\ldots\lambda_{21}}}. \tag{8.8}$$

Remaining probabilities are expressed as p_1:

$$\left.\begin{array}{l}
p_2=\frac{\lambda_{12}}{\lambda_{21}}p_1,\\
p_3=\frac{\lambda_{23}\lambda_{12}}{\lambda_{32}\lambda_{21}}p_1,\\
\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\\
p_k=\frac{\lambda_{k-1,\,k}\ldots\lambda_{12}}{\lambda_{k,\,k-1}\ldots\lambda_{21}}p_1,\\
\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\\
p_n=\frac{\lambda_{n-1,\,n}\ldots\lambda_{12}}{\lambda_{n,\,n-1}\ldots\lambda_{21}}p_1.
\end{array}\right\} \tag{8.9}$$

Thus, the problem of "death and multiplication" is solved in general form are found the maximum probabilities of states.

Example 2. To find the maximum probabilities of states for the process of death and multiplication whose graph/count is shown on Fig. 4.41.

Solution. On formulas (8.8) and (8.9) we have

$$p_1=\frac{1}{1+\frac{2}{3}+\frac{2\cdot 1}{3\cdot 2}+\frac{2\cdot 1\cdot 3}{3\cdot 2\cdot 2}}=\frac{1}{1+\frac{2}{3}+\frac{1}{3}+\frac{1}{2}}=\frac{2}{5},$$

$$p_2=\frac{2}{3}\cdot\frac{2}{5}=\frac{4}{15},\quad p_3=\frac{1}{3}\cdot\frac{2}{5}=\frac{2}{15},\quad p_4=\frac{1}{2}\cdot\frac{2}{5}=\frac{1}{5}.$$

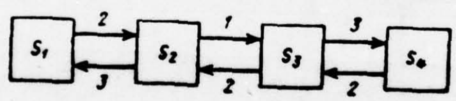


Fig 4.41

Page 226.

Example 3. Instrument consists of three assemblies; the flow of failures - simplest, the mean time of the failure-free operation of each assembly is equal to $\bar{t}_б$. The refused assembly immediately begins to be overhauled; the mean time of the repair (restoration/reduction) of assembly is equal to $\bar{t}_р$; the law of the distribution of this time exponential (flow of restoration/reductions - simplest). To find the average efficiency of instrument, if with three working assemblies it is equal to 100o/o, with two - 50o/o, and with one and less - an instrument not at all it works.

Solution. The enumeration of the states of system and the graph/count of states already were given in an example of 1 this paragraph. Let us label this graph/ccunt, i.e., let us write at each arrow/pointer appropriate intensity λ_{ij} (see Fig. 4.42).

Since the flow of the failures of each assembly - simplest, then time interval between failures in this flow is distributed according to exponential law with parameter $\lambda = 1/\bar{t}_6$, where $\bar{t}_6$— average the time of the failure-free operation of assembly.

On arrow/pointers to the right the system translates failures. If system is in state S_0, then work three assemblies; each of them undergoes the flow of failures with an intensity of $1/\bar{t}_6$; that means that the flow of failures, which functicns cn entire system, three times is more intensive: $\lambda_{01} = 3/\bar{t}_6$.

If system is in state S_1, then work two assemblies; the common/general/total flow cf failures has the intensity: $\lambda_{12} = 2/\bar{t}_6$ is analogous $\lambda_{23} = 1/\bar{t}_6$.

On arrow/pointers to the left the system translates repairs (restoration/reductions). Mean recovery time of assembly is equal to

$\bar{t}_p$ which means, that the intensity of flow of restoration/reductions, which functions on one restorable assembly, is equal to $\mu = 1/\bar{t}_p$, to two assembly - $2/\bar{t}_p$, to three assembly - $3/\bar{t}_p$. These values λ_{10}, λ_{21}, λ_{32} are written on Fig. 8.5 of the arrow/pointers, which lead to the left.

Using the obtained above general solution of the problem of death and multiplication, we have (placing p_0 instead of p_1):

$$p_0 = \frac{1}{1 + 3\left(\frac{\bar{t}_p}{\bar{t}_б}\right) + 3\left(\frac{\bar{t}_p}{\bar{t}_б}\right)^2 + \left(\frac{\bar{t}_p}{\bar{t}_б}\right)^3};$$

$$p_1 = 3\left(\frac{\bar{t}_p}{\bar{t}_б}\right) p_0;$$

$$p_2 = 3\left(\frac{\bar{t}_p}{\bar{t}_б}\right)^2 p_0;$$

$$p_3 = \left(\frac{\bar{t}_p}{\bar{t}_б}\right)^3 p_0.$$

Let us assign concrete/specific/actual values $\bar{t}_б = 10$ (hour), $\bar{t}_p = 5$ (hour). Then $\frac{\bar{t}_p}{\bar{t}_б} = 0{,}5$ and

$$p_0 = \frac{1}{1 + {}^3/_2 + {}^3/_4 + {}^1/_8} = {}^8/_{27},\ p_1 = {}^3/_2 \cdot {}^8/_{27} = {}^{12}/_{27},\ p_2 = {}^3/_4 \cdot {}^8/_{27} = {}^6/_{27},\ p_3 = {}^1/_8 \cdot {}^8/_{27} = {}^1/_{27}.$$

The average efficiency of instrument in the steady-state

conditions/mode:

$$100\%\,p_2 + 50\%\,p_1 = \left(\frac{800}{27} + \frac{600}{27}\right)\% = 51{,}9\%\ \text{rating.}$$

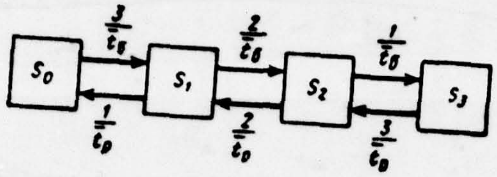


Fig. 4.42.

Page 227.

9. Cyclic process.

The Markovian process, which takes place in system, is called cyclic, if states are connected into ring (cycle) with the one-sided transitions (see Fig. 4.43 on page 228).

Let us write algebraic equations for the maximum probabilities of the states:

$$\left.\begin{array}{c} \lambda_{23} p_2 = \lambda_{12} p_1, \\ \lambda_{34} p_3 = \lambda_{23} p_2, \\ \dots\dots\dots\dots \\ \lambda_{k-1,\,k} p_{k-1} = \lambda_{k,\,k+1} p_k, \\ \dots\dots\dots\dots \\ \lambda_{n-1,\,n} p_{n-1} = \lambda_{n,\,1} p_n, \\ \lambda_{n,\,1} p_n = \lambda_{12} p_1, \end{array}\right\} \qquad (9.1)$$

plus the normalizing condition

$$p_1 + p_2 + \dots + p_n = 1.$$

From equations (9.1), after reject/throwing the latter, it is expressed all probabilities $p_2, \dots, p_n$ through p_1:

$$\left.\begin{array}{l}
p_2 = \dfrac{\lambda_{12}}{\lambda_{23}} p_1, \\
p_3 = \dfrac{\lambda_{23}}{\lambda_{34}} p_2 = \dfrac{\lambda_{12}\lambda_{23}}{\lambda_{23}\lambda_{34}} p_1 = \dfrac{\lambda_{12}}{\lambda_{34}} p_1, \\
p_4 = \dfrac{\lambda_{12}}{\lambda_{45}} p_1, \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
p_k = \dfrac{\lambda_{12}}{\lambda_{k,\, k+1}} p_1, \\
\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
p_n = \dfrac{\lambda_{12}}{\lambda_{n1}} p_1.
\end{array}\right\}$$

Page 228.

Substituting these expressions in (9.2), we will obtain:

$$p_1 + \lambda_{12}\left(\frac{1}{\lambda_{23}} + \frac{1}{\lambda_{34}} + \dots + \frac{1}{\lambda_{n1}}\right) p_1 = 1,$$

whence

$$\left.\begin{aligned} p_1 &= \frac{1}{1 + \lambda_{12}\left(\frac{1}{\lambda_{23}} + \frac{1}{\lambda_{34}} + \dots + \frac{1}{\lambda_{n1}}\right)}, \\ p_2 &= \frac{\lambda_{12}}{\lambda_{23}} p_1, \\ p_3 &= \frac{\lambda_{12}}{\lambda_{34}} p_1, \\ &\dots\dots\dots \\ p_k &= \frac{\lambda_{12}}{\lambda_{k,k+1}} p_1, \\ &\dots\dots\dots \\ p_n &= \frac{\lambda_{12}}{\lambda_{n1}} p_1. \end{aligned}\right| \qquad (9.2)$$

Formulas (9.2), which express the maximum probabilities of states for a cyclic process, can be led to more convenient and more demonstrative form, if we pass from intensities λ_{ij} to the mean times $\bar{t}_i$ of the stay of system (in a row) in state S_i (i = 1, ..., n).

It is real/actual, let from state S_i, as it takes place in

cyclic set-up, proceeds only one arrow/pointer (Fig. 4.44). Let system S is be in state S_i. Let us find the mathematical expectation of time T_i, which it still will stay in this state. Since process is Markov process, the law of time allocation of T_i does not depend on that, how long system already stayed in state S_i; that means it the same, which it would be, if system recently arrived into state S_i, i.e., represents by itself nothing else but the exponential law of the distribution of time interval T between adjacent events in the simplest "flow of the attendance/departures" of system from state S_i.

The parameter of this law is equal to $\lambda_{i,i+1}$, and the mean retention time of system in state S_i (if it in it already is located) equally to $\bar{t}_i = \frac{1}{\lambda_{i,i+1}}$. Hence $\lambda_{i,i+1} = \frac{1}{\bar{t}_i}$ for all i = 1, 2, ..., n - 1. For i = n, we will obtain (by the force of cyclic recurrence) $\lambda_{n,1} = \frac{1}{\bar{t}_n}$.

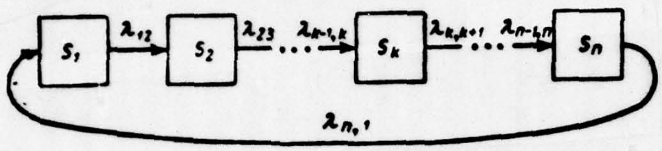


Fig. 4.43.

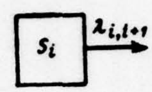


Fig. 4.44.

Page 229.

After substituting these expressions into formulas (9.2) after elementary transformations we will obtain:

$$p_1=\frac{\bar t_1}{\bar t_1+\bar t_2+\ldots+\bar t_n},$$
$$p_2=\frac{\bar t_2}{\bar t_1+\bar t_2+\ldots+\bar t_n},$$
$$\ldots\ldots\ldots\ldots$$
$$p_n=\frac{\bar t_n}{\bar t_1+\bar t_2+\ldots+\bar t_n},$$

or, are shorter:

$$p_k=\frac{\bar t_k}{\sum\limits_{i=1}^{n}\bar t_i}\quad(k=1,\ \ldots,\ n),\tag{9.3}$$

i.e. the maximum probabilities of states in the cyclic set-up belong as mean retention times of system in a row in each of the states.

Example 1. Electronic digital computer can be located in one of the following states:

S_1 - exact, it works.

S_2 — defective, stopped; is conducted the search of malfunction.

S_3 - malfunction it is localized; is conducted repair.

S_4 - repair it is finished; is conducted launch preparation of machine.

All flows of events - simplest. The mean time of the failure-free operation ETsVM (in a row) is equal to 0.5 (days). For a

repair machine it is necessary to stop on the average at 6 hours. The search of malfunction lasts on the average of 0.5 hours. After the termination of repair, the machine is prepared for launching/starting on the average 1 hour. To find the maximum probabilities of states.

Solution. The graph/count of states takes the form of cyclic set-up (Fig. 4.45).

Let us determine mean retention time ETsVM in a row in each state:

$$\bar{t}_1 = {}^1/_2, \quad \bar{t}_2 = {}^1/_{48}, \quad \bar{t}_3 = {}^1/_4, \quad \bar{t}_4 = {}^1/_{24} \text{ (days)},$$

whence, on formulas (9.3):

$$p_1 = \frac{{}^1/_2}{{}^1/_2 + {}^1/_{48} + {}^1/_4 + {}^1/_{24}} = {}^{24}/_{39}, \quad p_2 = {}^1/_{39}, \quad p_3 = {}^{12}/_{39}, \quad p_4 = {}^2/_{39},$$

or, in decimal fractions,

$$p_1 = 0{,}615; \quad p_2 = 0{,}026; \quad p_3 = 0{,}308; \quad p_4 = 0{,}051.$$

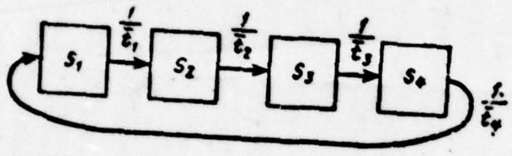


Fig. 4.45.

Page 230.

Thus, if process is reduced to simple cyclic with one-sided transitions, the maximum probabilities of states are located very simply: from the relationship/ratio of mean retention times (in a row) in each of the states.

In many instances of practice, it is necessary to deal with the branching cyclic process where the graph/count of states in separate units forms branchings off.

Example of 2. ETsVM can be located in the following states:

S_1 — exact, it works;

S_2 — defective, stopped; is conducted

S_3 —malfunction render/showed insignificant and it is removed by local resources;

S_4 - malfunction render/showed serious and he is removed by the brigade of the specialists;

S_5 - launch preparation.

Process, which takes place in system - Markov (all flows of events - simplest). The mean time of the exact work of machine is equal in a row $\bar{t}_1$, the mean retrieval time of malfunctions - $\bar{t}_2$, the mean time of repair by local resources - $\bar{t}_3$, the mean time of repair by the brigade of the specialists - $\bar{t}_4$, the mean time of preparation ETsVM for launching/starting - $\bar{t}_5$.

Malfunction ETsVM can be eliminated by local means with probability $\mathcal{P}$, and with probability $1 - \mathcal{P}$ requires the call of the brigade of the specialists. Brigade's work is paid in size/dimension of k (rubles/h).

It is required to find the maximum probabilities of states and to determine the average/mean expenditure/consumption, which goes for the payment of the work of maintenance crew per unit time (in a 24 hour period).

Solution. We construct the labeled graph/count of states (Fig. 4.46). If the state leaves only one arrow/pointer, then the intensity

of flow of events, which stands of this arrow/pointer, is equal to unity, divided into mean retention time (in a row) in this state. If the state leave not one arrow/pointer, but two, then the common/general/total intensity, equal to unity, divided into mean retention time (in a row) in this state, is multiplied for each arrow/pointer to probability that the transition will be completed precisely on this arrow/pointer.

equations for the maximum probabilities of states take the form:

$$\left.\begin{aligned} &\frac{1}{\bar{t}_1} p_1 = \frac{1}{\bar{t}_2} p_2, \\ &\frac{\mathcal{P}}{\bar{t}_2} p_2 = \frac{1}{\bar{t}_3} p_3, \\ &\frac{1-\mathcal{P}}{\bar{t}_2} p_2 = \frac{1}{\bar{t}_4} p_4, \\ &\frac{1}{\bar{t}_3} p_3 + \frac{1}{\bar{t}_4} p_4 = \frac{1}{\bar{t}_5} p_5, \\ &\frac{1}{\bar{t}_5} p_5 = \frac{1}{\bar{t}_1} p_1. \end{aligned}\right\} \tag{9.4}$$

plus the normalizing condition:

$$p_1 + p_2 + p_3 + p_4 + p_5 = 1. \tag{9.5}$$

Of equations (9.4) one as we know that it is possible to reject/throw; let us reject/throw the most complex - the fourth, while from the others is expressed p_2, p_3, p_4, p_5, through p_1:

$$p_2 = \frac{\bar{t}_2}{\bar{t}_1} p_1,$$

$$p_3 = \frac{\mathcal{P}\bar{t}_3}{\bar{t}_2} p_2 = \frac{\mathcal{P}\bar{t}_3}{\bar{t}_2} \frac{\bar{t}_2}{\bar{t}_1} p_1 = \frac{\mathcal{P}\bar{t}_3}{\bar{t}_1} p_1,$$

$$p_4 = \frac{(1-\mathcal{P})\,\bar{t}_4}{\bar{t}_2} p_2 = \frac{(1-\mathcal{P})\,\bar{t}_4}{\bar{t}_2} \frac{\bar{t}_2}{\bar{t}_1} p_1 = \frac{(1-\mathcal{P})\,\bar{t}_4}{\bar{t}_1},$$

$$p_5 = \frac{\bar{t}_5}{\bar{t}_1} p_1.$$

Page 231.

Substituting in (9.5), we have:

$$p_1\left(1 + \frac{\bar{t}_2}{\bar{t}_1} + \frac{\mathcal{P}\bar{t}_3}{\bar{t}_1} + \frac{(1-\mathcal{P})\,\bar{t}_4}{\bar{t}_1} + \frac{\bar{t}_5}{\bar{t}_1}\right) = 1.$$

Hence

$$p_1 = \frac{\bar{t}_1}{\bar{t}_1 + \bar{t}_2 + \mathcal{P}\bar{t}_3 + (1-\mathcal{P})\,\bar{t}_4 + \bar{t}_5},$$

$$p_2 = \frac{\bar{t}_2}{\bar{t}_1 + \bar{t}_2 + \mathcal{P}\bar{t}_3 + (1-\mathcal{P})\,\bar{t}_4 + \bar{t}_5},$$

$$p_3 = \frac{\mathcal{P}\bar{t}_3}{\bar{t}_1 + \bar{t}_2 + \mathcal{P}\bar{t}_3 + (1-\mathcal{P})\,\bar{t}_4 + \bar{t}_5},$$

$$p_4 = \frac{(1-\mathcal{P})\,\bar{t}_4}{\bar{t}_1 + \bar{t}_2 + \mathcal{P}\bar{t}_3 + (1-\mathcal{P})\,\bar{t}_4 + \bar{t}_5},$$

$$p_5 = \frac{\bar{t}_5}{\bar{t}_1 + \bar{t}_2 + \mathcal{P}\bar{t}_3 + (1-\mathcal{P})\,\bar{t}_4 + \bar{t}_5}.$$

The average fraction of time which the system carries out (in the steady-state conditions/mode) in state S_4 (repair by the brigade of the specialists) is equal to p_4. That means that per hour the system carries out in this state on the average p_4 hours. Multiplying this value on 24k, we will obtain the average expenditure of

resources for the payment of the brigade of the specialists for the days: $C = 24kp_4$.

Let us turn attention to the structures of probabilities p_1, p_2, ..., p_5 in the pattern of the branching cycle. They, so kA and in the case of simple cycle, represent by themselves the ratios of mean retention times (in a row) in states to the sum of all such times, with that difference, that for the state, which lies on "branch", what mean time is multiplied to transitional probability on this "branch" ($\mathcal{P}$ or $1 - \mathcal{P}$) Using this rule, it is possible to immediately write the maximum probabilities of states for left branching cyclic circuit.

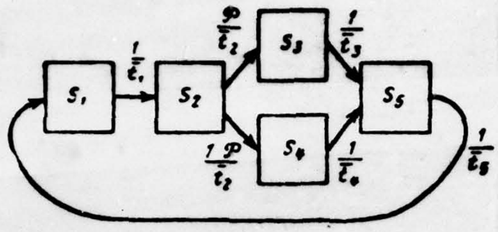


Fig. 4.46.

Page 232.

10. Approximate information of not- Markov processes to Markov. Method of "pseudostates".

In practice we almost never deal with Markov processes in the pure form: real processes almost always possess one or the other aftereffect. For a Markov process the retention time of system in any state is distributed in a row according to exponential law; in reality this hardly ever is thus. For example, if the flow of events, which translates system from state into state is a flow of the failures of some unit, then it is more logical to assume that the remaining time of the failure-free operation of unit depends on that, how long unit already worked. In this case, the retention time of unit in running order is random variable, distributed not on exponential, but according to some other law. Does arise the question

concerning that, is it possible to approximately substitute not-Poisson flows - Poisson and to which errors in the maximum probabilities of states can bring a similar replacement. For this, it is necessary to be able at least to approximately trace the random processes, which take place in systems with aftereffect.

Let us consider certain physical system S, in which proceeds the random process, directed by some not- Poisson flows of events. If we try for this process to write the equations, which express the probability states as of function of time, we will see, that in the general case this to us does not accomplish. It is real/actual, for a Markov system we computed probability that at torque/moment $t + \Delta t$ the system will be able S_i, taking into account only that, in which state system was at torque/moment t, and without taking into account that how long it was in this state. For a not-Markov system this method is already unsuitable: computing transitional probability from one state into another for time Δt, we must let us consider that how long system already led in this state. This is brought, instead of the ordinary differential equations, to equations with the partial derivatives, i. e., to the much more complex mathematical vehicle with the help of which only in rare cases it is possible to obtain necessary results.

Does arise the question: a it is not possible whether to reduce

artificially (at least approximately) a not- Markov process to Markov?

It turns out that in certain cases this is possible: namely, if the number of states of system not is very great, but differing from protozoa the flows of events, which participate in problem, represent by themselves (it is accurate or approximately) Erlang's flows. Then, introducing into the circuit of the possible states of system some fictitious "pseudostates", to reduce a not-Markov process to Markov and it is possible to describe it with the help of the ordinary differential equations which with $t \to \infty$ pass into algebraic equations for the maximum probabilities of states.

Let us explain the idea of the method of "pseudostates" based on specific example.

Page 233.

Example 1. Is examined S - technical equipment/device system, which can go out of order under the effect of the simplest flow of malfunctions with intensity λ. The refused equipment/device immediately begins to be restored. Recovery time (repair) T is distributed not according to the exponential law (as must so that the process will be Markov), but according to the law of Erlang of 3rd

order:

$$f_3(t) = \frac{\mu(\mu t)^2}{2} e^{-\mu t} \quad (t > 0). \qquad (10.1)$$

It is required to reduce this not- Markov process to Markov and to find for it the maximum probabilities of states.

Solution. Random variable T (recovery time) is distributed according to the law of Erlang and, which means, that it represents by itself the sum of three random variables T_1, T_2, T_3, distributed according to the exponential law (see § 5 chapters 4) with the parameter μ:

$$f_1(t) = \mu e^{-\mu t} \quad (t > 0). \qquad (10.2)$$

The true states of system a total of two:

S_1 - equipment/device is exact;

S_2 - equipment/device is restored.

The graph/count of these states is shown on Fig. 4.47 (is related to cyclic circuit).

However, considering that the transition on arrow/pointer $S_2 \to S_1$ takes place under the effect of not the simplest, and Erlang flow

of events, the process, which occurs in system, Markov is not, and for it we cannot write either differential, or algebraic equations.

In order to artificially reduce this process to Markov, let us introduce into the chain/network of states, instead of one state S_2, three consecutive "pseudostate".

$S_2^{(1)}$ - repair begins;

$S_2^{(2)}$ - repair is continued;

$S_2^{(3)}$ - repair is finished,

i.e. let us divide repair into three stage or "phase", moreover the retention time of system in each of the phases let us consider distributed according to exponential law (10.2). The graph/count of states will take the form, shown on Fig. 4.48, where the role of one state S_2 they will play three pseudostates $S_2^{(1)}$, $S_2^{(2)}$ and $S_2^{(3)}$. The process, which takes place in this system, will be already Markov.

Let us designate $p_2^{(1)}, p_2^{(2)}, p_2^{(3)}$ - maximum probabilities of the stay of system in pseudostates $S_2^{(1)}, S_2^{(2)}, S_2^{(3)}$; then

$$p_2 = p_2^{(1)} + p_2^{(2)} + p_2^{(3)}.$$

Designating

$$\bar{t}_1 = 1/\lambda, \quad \bar{t}_2 = 1/\mu,$$

we can immediately write (as for a usual cyclic circuit) maximum probability of the states:

$$\left.\begin{aligned} p_1 &= \frac{\bar{t}_1}{\bar{t}_1 + \bar{t}_2 + \bar{t}_2 + \bar{t}_2} = \frac{\bar{t}_1}{\bar{t}_1 + 3\bar{t}_2}; \\ p_2^{(1)} &= p_2^{(2)} = p_2^{(3)} = \frac{\bar{t}_2}{\bar{t}_1 + 3\bar{t}_2}; \\ p_2 &= p_2^{(1)} + p_2^{(2)} + p_2^{(3)} = \frac{3\bar{t}_2}{\bar{t}_1 + 3\bar{t}_2}. \end{aligned}\right\}$$

Page 234.

Let us note that value $3\bar{t}_2$ represents by itself nothing else but mean recovery time (repair) - it equal to the sum mean retention times of system in each phase of repair.

Passing in formulas for p_1, p_2 from the mean times $\bar{t}_1$, $\bar{t}_2$ to intensities of flow, on formulas $\bar{t}_1 = 1/\lambda$, $\bar{t}_2 = 1/\mu$, we will obtain:

$$p_1 = \mu/(\mu + 3\lambda), \quad p_2 = 3\lambda/(\mu + 3\lambda). \qquad (10.3)$$

Similarly is obtained the conclusion/derivation: for our elementary example the probability of the stay in each of two states, as for a Markov cycle, it is equal to relative mean retention time in a row in each of state.

A following example there will be somewhat more complex.

Example 2. The technical equipment/device S consists of two identical assemblies each of which can go out of order (reject) under the effect of the simplest flow of malfunctions with intensity λ. The refused node/unit immediately begins to be overhauled. The time of repair T is distributed according to the law of Erlang of second order:

$$f_2(t) = \mu^2 t e^{-\mu t} \quad (t > 0).$$

It is required to find the maximum probabilities of the states of system.

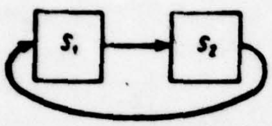
Solution. The true states of system three (we label them according to the number of refused node/units).

S_0 - both node/unit work;

S_1 - one node/unit works, another is overhauled;

S_2 - both node/unit are overhauled.

Let us divide conditionally repair into two phases: the repair is begun and repair is finished.



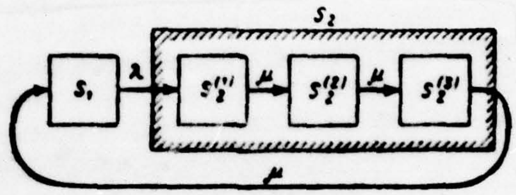


Fig. 4.47.

Fig. 4.48.

Page 235.

The duration of each phase let us consider distributed according to exponential law (10.2). The process, which occurs in system, is led to Markov, if we introduce these pseudostates:

$S_2^{(1)}$ - one node/unit works, another begins to be overhauled;

$S_2^{(2)}$ - one node/unit works, another ends to be overhauled;

$S_2^{(1,1)}$ - both node/unit begin to be repaired;

$S_2^{(1,2)}$ - one node/unit begin to be repaired;

$S_2^{(2,2)}$ - both node/unit end to be overhauled.

The graph/count of the states of system with pseudostates is shown on Fig. 4.49. On the arrow/pointers, which lead from $S_2^{(1,1)}$ in $S_2^{(1,2)}$ and from $S_2^{(2,2)}$ in $S_1^{(2)}$, it is written 2μ, but not μ, because to pass during the following phase of repair (termination of repair) can any of two node/units.

Equations for the maximum probabilities of states they take the form:

$$
\left.\begin{aligned}
&2\lambda p_0 + \mu p_2^{(1,2)} = (\lambda+\mu)\, p_1^{(1)},\\
&\mu p_1^{(1)} + 2\mu p_2^{(2,2)} = (\lambda+\mu)\, p_1^{(2)},\\
&\lambda p_1^{(1)} = 2\mu p_2^{(1,1)},\\
&2\mu p_2^{(1,1)} + \lambda p_1^{(2)} = 2\mu p_2^{(1,2)},\\
&\mu p_2^{(1,2)} = 2\mu p_2^{(2,2)},\\
&\mu p_1^{(2)} = 2\lambda p_0.
\end{aligned}\right| \qquad (10.4)
$$

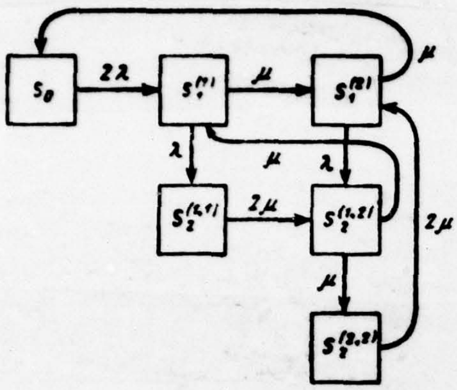


Fig. 4.49.

Page 236.

From third, fifth and sixth equations (10.4) we have:

$$\left.\begin{aligned} p_2^{(1,1)} &= \frac{\lambda}{2\mu}\, p_1^{(1)}, \\ p_2^{(2,2)} &= \frac{1}{2}\, p_2^{(1,2)}, \\ p_1^{(2)} &= \frac{2\lambda}{\mu}\, p_0. \end{aligned}\right| \qquad (10.5)$$

that it makes it possible tc decrease the number of unknowns: substituting (10.5) in remaining three equations (10.4), we will obtain:

$$\left.\begin{aligned} 2\lambda p_0 + \mu p_2^{(1,2)} &= (\lambda+\mu)\, p_1^{(1)}, \\ \mu p_1^{(1)} + \mu p_2^{(1,2)} &= (\lambda+\mu)\, \frac{2\lambda}{\mu}\, p_0, \\ \lambda p_1^{(1)} + \frac{2\lambda^2}{\mu}\, p_0 &= 2\mu p_2^{(1,2)}. \end{aligned}\right| \qquad (10.6)$$

From these three equations with three unknowns p_0, $p_1^{(1)}$, $p_2^{(1,2)}$ it is possible on arbitrariness to reject/throw any, for example, latter, and to supplement the normalizing condition:

$$p_0 + p_1^{(1)} + p_1^{(2)} + p_2^{(1,1)} + p_2^{(1,2)} + p_2^{(2,2)} = 1,$$

or, taking into account (10.5),

$$\left(1 + \frac{2\lambda}{\mu}\right) p_0 + \left(1 + \frac{\lambda}{2\mu}\right) p_1^{(1)} + \frac{3}{2}\, p_2^{(1,2)} = 1. \qquad (10.7)$$

It is solved two first equation (10.6) together with equation (10.7). Is expressed from first equation $p_2^{(1,2)}$ through p_0 and $p_1^{(1)}$:

$$p_2^{(1,2)} = \left(1 + \frac{\lambda}{\mu}\right) p_1^{(1)} - \frac{2\lambda}{\mu} p_0 \qquad (10.8)$$

let us substitute this expression in the second equation; we will obtain:

$$(\lambda + 2\mu)\, p_1^{(1)} = \left[2\lambda + \frac{2\lambda}{\mu}(\lambda + \mu)\right] p_0,$$

or, after reduction on $(\lambda + 2\mu)$:

$$p_1^{(1)} = \frac{2\lambda}{\mu} p_0. \qquad (10.9)$$

Substituting this in (10.8), it is expressed and probability $p_2^{(1,2)}$ through p_0:

$$p_2^{(1,2)} = \frac{2\lambda^2}{\mu^2} p_0. \qquad (10.10)$$

By now let us substitute (10.9) and (10.10) under normalizing condition (10.7):

$$p_0\left(1 + \frac{2\lambda}{\mu} + \frac{2\lambda}{\mu} + \frac{\lambda^2}{\mu^2} + \frac{3\lambda^2}{\mu^2}\right) = 1,$$

whence

$$p_0 = \frac{1}{1 + 4\lambda/\mu + 4\lambda^2/\mu^2} = \frac{\mu^2}{\mu^2 + 4\lambda\mu + 4\lambda^2}. \qquad (10.11)$$

After this let us find all the remaining maximum probabilities: from (10.9), (10.10)

$$p_1^{(1)} = \frac{2\lambda\mu}{\mu^2 + 4\lambda\mu + 4\lambda^2}, \quad p_2^{(1,2)} = \frac{2\lambda^2}{\mu^2 + 4\lambda\mu + 4\lambda^2};$$

from (10.5):

$$p_1^{(2)} = \frac{2\lambda\mu}{\mu^2 + 4\lambda\mu + 4\lambda^2}, \quad p_2^{(1,1)} = \frac{\lambda^2}{\mu^2 + 4\lambda\mu + 4\lambda^2},$$
$$p_2^{(2,2)} = \frac{\lambda^2}{\mu^2 + 4\lambda\mu + 4\lambda^2}$$

Page 237.

After are found the probabilities of pseudostates, it is possible to find the probabilities of the states:

$$p_0 = \frac{\mu^2}{\mu^2 + 4\lambda\mu + 4\lambda^2}; \quad p_1 = p_1^{(1)} + p_1^{(2)} = \frac{4\lambda\mu}{\mu^2 + 4\lambda\mu + 4\lambda^2};$$
$$p_2 = p_2^{(1,1)} + p_2^{(1,2)} + p_2^{(2,2)} = \frac{4\lambda^2}{\mu^2 + 4\lambda\mu + 4\lambda^2}.$$

for example, when $\lambda = 1$, $\mu = 4$ (in steady state) probability that both node/unit they work, it is equal to $p_0 = 16/25 = 0.64$; probability that one node/unit is overhauled $p_1 = 8/25 = 0.32$; probability that both node/unit are overhauled $p_2 = 1/25 = 0.04$.

Let us note that the method of pseudostates admits comparatively simple solution of problem only in the simplest cases when the number of states of reference system is small. However, sometimes it is possible to use this method, also, to the problems where the number of states not is very small; in any case, to obtain if not literal, then numerical approximate solution of the matching system of linear algebraic equations.

The possibilities of the method of pseudostates substantially are widened, if we use as the flows of events not some Erlang flows alone in pure form, but also by the generalized Erlang and mixed generalized Erlang distributions which it was mentioned at the end §5.